

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 22, 2002, 07:37:11 ; Search time 157.11 Seconds  
(without alignments)  
8332.605 Million cell updates/sec

Title: US-09-604-231-1  
Perfect score: 1527  
Sequence: 1 ctcatgcatctgcgcgtt.....gttgaaaccttgatgttcg 1527

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_1101.\*  
1: /SID52/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SID52/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SID52/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SID52/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SID52/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SID52/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SID52/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SID52/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SID52/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SID52/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SID52/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SID52/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SID52/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SID52/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SID52/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SID52/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SID52/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SID52/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SID52/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SID52/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SID52/gcgdata/geneseq/geneseq/NA2000.DAT.\*  
22: /SID52/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1527	100.0	1527	22	AAF31528
2	1315	99.2	34980	22	AAH68533
3	1492	97.7	1983	22	AAH67869
4	1412.6	92.5	5969	22	AAF32543
5	1107.4	72.5	1109	22	AAF31529
6	190.6	12.5	2913	22	AAH54100
7	151.2	9.9	5840	20	AAH12968
8	126.6	8.3	3895	19	AAV52334
9	108	7.1	3615	15	AAQ55752
10	107.8	7.1	9769	19	AAV52163
11	97.8	6.4	465	18	AAAT47502

12	86.2	5.6	30246	18	AAV74367	Staphylococcus aur
13	76.4	5.0	465	18	AAV75628	Staphylococcus aur
14	74.2	4.9	474	20	AAH13576	Enterococcus faeca
15	71.6	4.7	357	18	AAAT47505	Partial Enterobact
16	70.4	4.6	9797	20	AAH13487	Enterococcus faeca
17	67	4.4	8494	19	AAH52296	Streptococcus pneu
18	66	4.3	1947	22	AAH81338	Escherichia coli p
19	60.4	4.0	2550	22	AAH54979	S. epidermidis gen
20	60	3.9	6285	20	AAH13352	Enterococcus faeca
21	59.2	3.9	2049	22	AAH68426	C. glutamicum codin
22	59.2	3.9	2172	22	AAH13542	C. glutamicum phosp
23	59.2	3.9	34980	22	AAH68528	C. glutamicum codin
24	56.8	3.7	29555	18	AAV74517	Staphylococcus aur
25	54.2	3.5	428	22	AAF31544	C. glutamicum phosp
26	53.2	3.5	2996	22	AAH54445	S. epidermidis gen
27	53.2	3.5	3081	22	AAH54946	S. epidermidis gen
28	53.2	3.5	3932	22	AAH54056	S. epidermidis gen
29	51.8	3.4	7156	20	AAH12966	Enterococcus faeca
30	51.4	3.4	1906	20	AAH13595	Enterococcus faeca
31	51.4	3.4	2882	19	AAV52273	Streptococcus pneu
32	51	3.3	29555	18	AAV74517	Staphylococcus aur
33	49.2	3.2	796	18	AAV74736	Staphylococcus aur
34	45	2.9	3604	20	AAH13417	Enterococcus faeca
35	44.6	2.9	7900	18	AAV74449	Staphylococcus aur
36	44.2	2.9	6092	20	AAH13162	Enterococcus faeca
37	43.2	2.8	760	19	AAZ96397	S. pneumoniae deri
38	42	2.8	400	18	AAV78285	Staphylococcus aur
39	41.6	2.7	341	18	AAV78546	Staphylococcus aur
40	41.4	2.7	10732	21	AAH10594	Gene encoding a su
41	41.2	2.7	6590	20	AAH13048	Enterococcus faeca
42	40.6	2.7	567	21	AAZ29550	HIV codon altered
43	40	2.6	249	20	AAH13801	Enterococcus faeca
44	39.8	2.6	1314	18	AAH84083	DNA encoding a S.
45	39.8	2.6	1314	19	AAV53453	DNA encoding a Pts

## ALIGNMENTS

RESULT 1  
AAF31528  
ID AAF31528 standard; DNA; 1527 BP.  
XX  
AC AAF31528;  
XX  
DT 09-APR-2001 (first entry)  
XX  
DE C. glutamicum phosphoenolpyruvate DNA #1.  
XX  
KW Phosphoenolpyruvate; sugar phosphotransferase system; PTS; ds.  
XX  
OS Corynebacterium glutamicum.  
XX  
WO200102583-A2.  
XX  
PD 11-JAN-2001.  
XX  
PF 27-JUN-2000; 2000WO-IB00973.  
XX  
PR 01-JUL-1999; 99US-0142691.  
PR 23-AUG-1999; 99US-0150310.  
PR 03-SEP-1999; 99DE-1042095.  
PR 03-SEP-1999; 99DE-1042097.  
XX  
(BADI ) BASF AG.  
XX  
Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;  
XX  
WPI; 2001-080989/09.  
XX  
Corynebacterium glutamicum nucleic acids encoding phosphoenolpyruvate;  
XX  
sugar phosphotransferase system proteins or their portions, useful for  
PT  
typing or identifying C. glutamicum or related bacteria, and as markers

PT for transformation -

Claim 3; Page 98-101; 144pp; English.

The present invention relates to *Corynebacterium glutamicum* phosphoenolpyruvate: sugar phosphotransferase system (PTS) proteins. The PTS nucleic acids and proteins are useful in the identification of microorganisms which can be used to produce fine chemicals, for modulating fine chemical production in *C. glutamicum* or related bacteria, the typing or identification of *C. glutamicum* or related bacteria, as reference points for mapping *C. glutamicum* genome, and as markers for transformation.

SQ Sequence 1527 BP; 304 A; 392 C; 430 G; 401 T; 0 other;

```
Query Match      100.0%; Score 1527; DB 22; Length 1527;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1527; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Qy	1	ctaatgcatctgcgcgcgttccgcgttctcttccagtgttggtccagtgttggtgtttccacgcgaaccaag	60
Db	1	ctaatgcatctgcgcgcgttccgcgttctcttccagtgttggtccagtgttggtgtttccacgcgaaccaag	60
Qy	61	cgtttcgcgcgcaatgaattcttcctgggcgcgcgtatttggtatggcgcatgggtgttcccgag	120
Db	61	cgtttcgcgcgcaatgaattcttcctgggcgcgcgtatttggtatggcgcatgggtgttcccgag	120
Qy	121	cttggtgaacgcgtctacgacgtggccgcacacatggctgcgggcgaaatgccaaatgtggc	180
Db	121	cttggtgaacgcgtctacgacgtggccgcacacatggctgcgggcgaaatgccaaatgtggc	180
Qy	181	ctgtttggtttagatgtttgcccaagcccggttaccaggcgacccgttcttctgtgctggt	240
Db	181	ctgtttggtttagatgtttgcccaagcccggttaccaggcgacccgttcttctgtgctggt	240
Qy	241	ggttctctgattctgcgaacgacagaaagtccctgcacaaagcactcaagggaactgc	300
Db	241	ggttctctgattctgcgaacgacagaaagtccctgcacaaagcactcaagggaactgc	300
Qy	301	agacttctgatactctccagtgtgcgttgcgtctcacccggattcttaccttcac	360
Db	301	agacttctgatactctccagtgtgcgttgcgtctcacccggattcttaccttcac	360
Qy	361	cattggcccgaatcgctctggtggcgcatgtctggcacacaggtctcacagggaacttta	420
Db	361	cattggcccgaatcgctctggtggcgcatgtctggcacacaggtctcacagggaacttta	420
Qy	421	tgatttcggttggtccagtcggcggtctgctcttcggtctggtctactcaccaatcgctat	480
Db	421	tgatttcggttggtccagtcggcggtctgctcttcggtctggtctactcaccaatcgctat	480
Qy	481	cactggtctgcacacagtccttccgcgcaatttgagctggagctgtttaaaccagggtggatc	540
Db	481	cactggtctgcacacagtccttccgcgcaatttgagctggagctgtttaaaccagggtggatc	540
Qy	541	cttcatcttcgcaacgcatctatgctataatcgcgccagggtgcgcgatgtttggcagt	600
Db	541	cttcatcttcgcaacgcatctatgctataatcgcgccagggtgcgcgatgtttggcagt	600
Qy	601	gttctcttcggcgaagagtgaagctcaaggcccttgcaagctgcaggtgcgtcaggtgtctccgc	660
Db	601	gttctcttcggcgaagagtgaagctcaaggcccttgcaagctgcaggtgtctcaggtgtctccgc	660
Qy	661	tgttcttggtattacggagcctgcgacatttcgggtgtgaaccttcgctgcgtgcgcgtt	720
Db	661	tgttcttggtattacggagcctgcgacatttcgggtgtgaaccttcgctgcgtgcgcgtt	720
Qy	721	cttcatcggtatcggtaccgcgcatcggtggcgctttgattgactctttaaatacaaa	780
Db	721	cttcatcggtatcggtaccgcgcatcggtggcgctttgattgactctttaaatacaaa	780
Qy	781	ggcagttgcgttgggcgcgtcagcaggtttcttgggtgtgtttctatttgaagctcccaaat	840









```
QY 1021 caccatccaggacaccttgaccggtgaagctattgcactgagcagcgtcagcgatgc 1080
    |||||
Db 5278 caccatccaggacaccttgaccggtgaagctattgcactgagcagcgtcagcgatgc 5337
    |||||
QY 1081 catgttccagcggaagattggctggcgggtggccatcgctcccaaccaggggcagtt 1140
    |||||
Db 5338 catgttccagcggaagattggctggcgggtggccatcgctcccaaccaggggcagct 5397
    |||||
QY 1141 agttctccggtgagtgaaagattggtggcattcccaatcgccatcgctttcgcagt 1200
    |||||
Db 5398 ggtttccaccagtgagcggaagatcggtggccttcccatctggtcacgctttcgcagt 5457
    |||||
QY 1201 tcgcaccaagctgagatggttccaatgtggatatcttgatgcacattggttttcgcac 1260
    |||||
Db 5458 ccgcactaaagctgagatggttccaatgtggatatcttgatgcacattggttttcgcac 5517
    |||||
QY 1261 agtaaacctcaacggcagcagcactttaaccgcgtgaagaagcaggcgcatgaagtcaaac 1320
    |||||
Db 5518 cgtaaacctcaacggcagcagcactttaaccgcgtgaagaagcaggcgcatgaagtcaaac 5577
    |||||
QY 1321 agggagagctggtgaaattcgatattgatgccattaaagctgcaaggttatgaagtaac 1380
    |||||
Db 5578 agggagagctggtgaaattcgatattgatgccattaaagctgcaaggttatgaagtaac 5637
    |||||
QY 1381 cagccgattgttttcgaattacaagaaacccggacctgtaaacacttacggtttggg 1440
    |||||
Db 5638 cagccgattgttttcgaattacaagaaacccggacctgtaaacacttacggtttggg 5697
    |||||
QY 1441 cgaattgaagcgggagcgaacctgtcaacgtcgcaagaagaagcgggtgccagcaac 1500
    |||||
Db 5698 cgaattgaagcgggagcgaacctgtcaacgtcgcaagaagaagcgggtgccagcaac 5757
    |||||
QY 1501 accataagttgaaccttgatgttcg 1527
    |||||
Db 5758 accataagttgaaccttgatgttcg 5784
    |||||

RESULT 5
AAF31529
ID AAF31529 standard; DNA; 1109 BP.
XX
AC AAF31529;
XX
XX 09-APR-2001 (first entry)
DE C.glutamicum phosphoenolpyruvate DNA #2.
XX
KW Phosphoenolpyruvate; sugar phosphotransferase system; PTS; ds.
XX
OS Corynebacterium glutamicum.
XX
PN WO200102583-A2.
XX
PD 11-JAN-2001.
XX
XX 27-JUN-2000; 2000WO-IB00973.
XX
PR 01-JUL-1999; 99US-0142691.
PR 23-AUG-1999; 99US-0150310.
PR 03-SEP-1999; 99DE-1042095.
PR 03-SEP-1999; 99DE-1042097.
XX
XX (BADI ) BASF AG.
XX
XX Pompejus M, Kroegeer B, Schroeder H, Zelder O, Haberhauer G;
PI WPI; 2001-080989/09.
XX
XX
XX Corynebacterium glutamicum nucleic acids encoding phosphoenolpyruvate:
PT sugar phosphotransferase system proteins or their portions, useful for
PT typing or identifying C. glutamicum or related bacteria, and as markers
PT for transformation -
```

```
XX
PS Claim 5; Page 103-104; 144pp; English.
XX
CC The present invention relates to Corynebacterium glutamicum
CC phosphoenolpyruvate: sugar phosphotransferase system (PTS) proteins.
CC The PTS nucleic acids and proteins are useful in the
CC identification of microorganisms which can be used to produce fine
CC chemicals, for modulating fine chemical production in C. glutamicum or
CC related bacteria, the typing or identification of C. glutamicum or
CC related bacteria, as reference points for mapping C. glutamicum genome,
CC and as markers for transformation.
XX
SQ Sequence 1109 BP; 236 A; 278 C; 303 G; 292 T; 0 other;

Query Match 72.5%; Score 1107.4; DB 22; Length 1109;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1108; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 419 tatgatttcggtggtccagtcggtcgtctcttcggtctggtctactcaccatcgtc 478
    |||||
Db 1 tatgatttcggtggtccagtcggtcgtctcttcggtctggtctactcaccatcgtc 60
    |||||
QY 479 atcaactggtctgcaccagtccttcgcccaattgagctggagctgtttaaccagggtgga 538
    |||||
Db 61 atcaactggtctgcaccagtccttcgcccaattgagctggagctgtttaaccagggtgga 120
    |||||
QY 539 tccttcattctgcaccagtccttcgcccaattgagctggagctgtttaaccagggtgga 598
    |||||
Db 121 tccttcattctgcaccagtccttcgcccaattgagctggagctgtttaaccagggtgga 180
    |||||
QY 599 gtgtcttcctgcggaagagtgaaagctcaaggcccttcgagtgcttcagggtctcc 658
    |||||
Db 181 gtgtcttcctgcggaagagtgaaagctcaaggcccttcgagtgcttcagggtctcc 240
    |||||
QY 659 gctgttcttggtattacggagcgtcgatcttcggtgagccttcgctgcgtgcgcg 718
    |||||
Db 241 gctgttcttggtattacggagcgtcgatcttcggtgagccttcgctgcgtgcgcg 300
    |||||
QY 719 ttcttcacgtggtatcggtacccagctatcggtggcgtttggttcattatgctccagat 778
    |||||
Db 301 ttcttcacgtggtatcggtacccagctatcggtggcgtttggttcattatgctccagat 360
    |||||
QY 779 aaggcagtgctgtggcgctgcaggtttcttggtggtgtgtttctattgctccagat 838
    |||||
Db 361 aaggcagtgctgtggcgctgcaggtttcttggtggtgtgtttctattgctccagat 420
    |||||
QY 839 atggtcatgttctgtgtgctgagttgttaccttcttcacgtcgtgcgcgcagatt 898
    |||||
Db 421 atggtcatgttctgtgtgctgagttgttaccttcttcacgtcgtgcgcgcagatt 480
    |||||
QY 899 gcttatggccttacttgggttcgccaacggcagcagattgacatcagatgcaacgcgtgct 958
    |||||
Db 481 gcttatggccttacttgggttcgccaacggcagcagattgacatcagatgcaacgcgtgct 540
    |||||
QY 959 ccagtgctgcaggaaacgacaaagccgaagcagaagcaccgcagaaatttcaaacgat 1018
    |||||
Db 541 ccagtgctgcaggaaacgacaaagccgaagcagaagcaccgcagaaatttcaaacgat 600
    |||||
QY 1019 tccaccatcatccaggcacctttgaccggtggaagctattgacctgagcagcgtcagcag 1078
    |||||
Db 601 tccaccatcatccaggcacctttgaccggtggaagctattgacctgagcagcgtcagcag 660
    |||||
QY 1079 gccatgtttgcaggcggaagcgttgcctcgggcggtgcccattgcccaaccaaggggcag 1138
    |||||
Db 661 gccatgtttgcaggcggaagcgttgcctcgggcggtgcccattgcccaaccaaggggcag 720
    |||||
QY 1139 ttagtttctccggtgagtggaagattgtgtggcattcccatctgcccatgctttcgca 1198
    |||||
Db 721 ttagtttctccggtgagtggaagattgtgtggcattcccatctgcccatgctttcgca 780
    |||||
QY 1199 gttcgcaccaagcgtcaggatggttccaatgtggatatcttgatgcacattggtttcgac 1258
    |||||
```

Db 781 gttcgaccaaggtgaggtgttccaatgtggatatcttgatgacattgtgttcgac 840  
 QY 1259 acagtaaaactcaacgcgcgaactttaaccgcgtgaagaagcagggcgatgaagtcataa 1318  
 Db 841 acagtaaaactcaacgcgcgaactttaaccgcgtgaagaagcagggcgatgaagtcataa 900  
 QY 1319 gcaaggagctgtgtgaattcgatattgatgccattaaagctgcaggttatgaggtta 1378  
 Db 901 gcaaggagctgtgtgaattcgatattgatgccattaaagctgcaggttatgaggtta 960  
 QY 1379 accacgccgattgttttcgaattacaagaacacggaccctgtaaacacttacggtttg 1438  
 Db 961 accacgccgattgttttcgaattacaagaacacggaccctgtaaacacttacggtttg 1020  
 QY 1439 ggcgaataatgaagcggagcccaacctgtcaacgtgcgaagaagaagcggcgccagca 1498  
 Db 1021 ggcgaataatgaagcggagcccaacctgtcaacgtgcgaagaagaagcggcgccagca 1080  
 QY 1499 acaccataagttgaaccttgagttgctg 1527  
 Db 1081 acaccataagttgaaccttgagttgctg 1109

## RESULT 6

AAH54100/c  
 ID AAH54100 standard; DNA; 2913 BP.

AC AAH54100;

XX 03-SEP-2001 (first entry)

DT S. epidermidis genomic polynucleotide sequence SEQ ID NO:3464.

DE Staphylococcus epidermidis SRI strain; Infection; diagnosis;

KW vaccination; endocarditis; ds.

XX Staphylococcus epidermidis.

OS WO200134809-A2.

XX 17-MAY-2001.

XX 09-NOV-2000; 2000WO-US30782.

XX 09-NOV-1999; 99US-0164258.

XX (GLAX ) GLAXO GROUP LTD.

XX Kimmerly WJ;

XX WPI; 2001-316495/33.

PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,  
 useful for vaccinating against infections, e.g. endocarditis -

PS Claim 8; Page 1030-1031; 2188pp; English.

XX AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides  
 CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.  
 CC (I) and (II) can have antibacterial activity and therefore can be used  
 CC in vaccination. The nucleic acids (I) may be used to produce the  
 CC S. epidermidis polypeptides (II) via the production of vectors  
 CC containing them which are used to produce hosts cells which express the  
 CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be  
 CC used to vaccinate subjects and to raise antibodies against the bacteria.  
 CC The polypeptides may also be used to assay for other inhibitors of their  
 CC activity and therefore identify compounds that may be used for the  
 CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to  
 CC AAH5090 represent specifically claimed S. epidermidis genomic DNA  
 CC polynucleotide sequences from the present invention. AAH5091 to  
 CC AAH5098 represent oligonucleotide sequences and primers which are used  
 CC in the exemplification of the present invention.  
 CC N.B. The present invention specifically claims all the polynucleotide

CC sequences given in the sequence listing of the present specification,  
 CC however the sequence listing only goes up to SEQ ID NO:4454 so even  
 CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,  
 CC no sequences are present for SEQ ID NO:4455 to 4464.

SQ Sequence 2913 BP; 1072 A; 480 C; 419 G; 942 T; 0 other;

Query Match 12.5%; Score 190.6; DB 22; Length 2913;  
 Best Local Similarity 55.0%; Pred. No. 2.7e-46;  
 Matches 468; Conservative 0; Mismatches 364; Indels 19; Gaps 4;

QY 5 tggcatctgcgcgttccttcctccagtggttggttttcacgcgaacacacgcgtt 64  
 Db 2690 TTGCAATGACACCTTTTACATTTATACCAATACTTATGTTGTTTGTAGTCAGCTAAGCGAT 2631  
 QY 65 tcggcgccaatgagttccctcctggcgccgcgtattggtatgctgcatggttcccgagcttg 124  
 Db 2630 TTGGTGGTAATCCTTTATTTAGGTGCAGC-TCTAGGTATGATACTTGTTCATCCTGGATTG 2572  
 QY 125 gtaacggctacgaagcgtggcgccaccatggctgcggggcaa---atgccaatgtggtcc 181  
 Db 2571 ATGAGTGCATATGATTTTCCAAAAGCTTTAGAAAGAGGAAAGCTATTCACACTGGGAT 2512  
 QY 182 ctgtttggttttagatgttgcccaagccggttaccaggccaccgctctcctgtgctgggtg 241  
 Db 2511 GTCCTTTGGGCTACATATTAATGAAGTAGGTTATCAGGGACAAAGTATACCTATGCTCGTA 2452  
 QY 242 gttcttgattctggcaacgatcgagaagttcctgcacaagcgcactcaaggcgactgca 301  
 Db 2451 GCAACATATATTTAGCTACGATTTGAAAATGTTACGTAAAGTTATTCCAACTGTGTTA 2392  
 QY 302 gacttctgatcactccagtgctgacgttgctgctcaccgggattccttaacatcagcc 361  
 Db 2391 GATAAATTTATTGACGCCATTTATCAATTTTATTACAGCATTTATAACATTTTATTTT 2332  
 QY 362 attgcccagcaatgcgctgggtggcgatgctggcacacggtctcacaggacttat 421  
 Db 2331 GTAGGGCCTGTCACTCGTCAATAGGTTATTTGGTTATCTCATGATGATGACTTGGTTATAT 2272  
 QY 422 gattcgggtggtccagtcgcggtctgctctcctcgtgctgctactcaaccaatcgtcatc 481  
 Db 2271 GAATTTGGTGGAGCTATTTGGTGGACTTATATTTGGTTTATTATATATGCGCCAATCGTCATT 2212  
 QY 482 actggtctgcaccagctccctcccgccaattgagctg-----gagctgtttaac 529  
 Db 2211 ACAGGAATGCACCATAGCTTTATTGCAAGTTGAAACGACATTAATAATGCTGATCGACTAAA 2152  
 QY 530 cagggtggatccctcatcttcgcaacggcatctatggtaaatatcgccagggtgcggca 589  
 Db 2151 ACAGGTGGTTCATTTATCTTCCCAATCGCAACGATGTCAAAATATTGCACAGGTGGTGCA 2092  
 QY 590 tgtttggcagtgcttct---tcctggcgaagagtgaagaagctcaaggcccttcgagtgct 646  
 Db 2091 GCTTTAGCTGCATTTCTTTATCATTAAGCAAAATAAAAAATTAAGAGGTGTGCTTCGCG 2032  
 QY 647 tcagggtctccgcttcttctgattatcaggagcctgcgcatcttcggtgtgaaaccttcgc 706  
 Db 2031 GCGGGTATTTCAGCTTTTACTAGGAATTACAGAACCCAGCAATGTTTGGTGTCAATCTAAA 1972  
 QY 707 ctgctgtgcccgttcttctcatcgggtatcggtacccgcagctatcggtggtgcgttttgatgca 766  
 Db 1971 TTGAGATATCCATTTATAGTGTGCTGTTCAGGATCAGGTATAGTGTGCGGCTTATATTCA 1912  
 QY 767 ctctttaatatacgaagcagttgctgtggcgctgcaggtttcttgggtgtgtttctatt 826  
 Db 1911 TTCTTCAAGTAAACGATAGCGCTTGGTACAGCTGGATTACCTGGATTATATCTATA 1852  
 QY 827 gatgtccacga 837  
 Db 1851 AATCCTACACA 1841

```

RESULT 7
AA12968
ID AAX12968 standard; DNA; 5840 BP.
XX
AC AAX12968;
XX
DT 19-MAR-1999 (first entry)
XX
XX Enterococcus faecalis genome contig SEQ ID NO:31.
DE DE
XX
XX Enterococcus faecalis; contig; detection: Enterococcal infection;
KW vaccine; attenuation; computer readable medium; ds.
XX
XX Enterococcus faecalis.
OS
XX
XX WO9850555-A2.
PN
XX
XX 12-NOV-1998.
PD
XX
XX 04-MAY-1998; 98WO-US08985.
PF
XX
XX 14-NOV-1997; 97US-0066009.
PR
XX 06-MAY-1997; 97US-0044031.
PR
XX 16-MAY-1997; 97US-0046655.
PR
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX
XX Barash SC, Dillon PJ, Kunsch CA;
PI
XX
XX WPI; 1999-045171/04.
DR
XX
XX New isolated Enterococcus faecalis polynucleotides and polypeptides
PT - used to develop products for the detection of Enterococcus and for
PT use in vaccines for prevention or attenuation of Enterococcus
PT infection.
XX
XX Claim 1; Page 389-392; 2084pp; English.
PS
XX
XX A computer readable medium has been developed which has recorded on it
CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.
CC AAX12938 to AAX13919 represent these nucleotide sequences which are
CC primary nucleotide sequences, also known as contigs. The computer-based
CC system can identify fragments of the Enterococcus faecalis genome with
CC commercial importance. The products can be used to detect the presence
CC of Enterococcus faecalis in samples. They can also be used for
CC diagnosing Enterococcal infection in an animal and monitoring
CC progression of disease, and for identifying agents which can be used to
CC modulate the growth or pathogenicity of Enterococcus faecalis, or
CC another related organism, in vivo or in vitro. In particular the
CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences
CC can be used in vaccines to prevent or attenuate an Enterococcal
CC infection.
XX
XX Sequence 5840 BP; 1848 A; 944 C; 1256 G; 1779 T; 13 other;
SQ

Query Match 9.9%; Score 151.2; DB 20; Length 5840;
Best Local Similarity 48.5%; Pred. No. 2,2e-34;
Matches 525; Conservative 0; Mismatches 533; Indels 24; Gaps 3;

QY 329 ttgctgtcaccggattcttaccatcgcattgcccagcaaatgcgtgggtgggc 388
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1850 ttcttattatggcactgttaacctcttagctattggtctcctcggacgggtcatcggc 1909
QY 389 gatgtctggcacacggtctacaggacattatgattcgggtggtccagtcggcggtctg 448
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1910 gatgtgtgggacaaggatatacaacgcaattacgctcttagtccaattatcgcagggtta 1969
QY 449 ctcttcggtctgtctactaccaaatcgtcatcactggtctgcaccagtcctcccgcca 508
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1970 ttaatgggtctgtgtggcaagctcttggaattgttgggtatgcattggggctttgacca 2029
QY 509 attgagctggagctgtttaaccagggtggatcctctatctcttcgcaacggcatctatggct 568

```

```

RESULT 8
AAV52334/c
ID AAV52334 standard; DNA; 3895 BP.
XX
AC AAV52334;
XX
DT 23-OCT-1998 (first entry)
XX
DE Streptococcus pneumoniae genome fragment SEQ ID NO:201.

```

```

QY 1406 aa 1407
DB ||
DB 2909 aa 2910

```

```

Db 2030 attatgatgttaaaacttaacacaagggtggcgatcacgattggtaccgatgtttattaccagcc 2089
QY 569 aatatgccccaggggtgcgcatgttttgccagtgcttctcctgcggaagagtgaaaaagctc 628
Db 2090 gttattgcaacaaggcgggtgcttttagctgtcttttcttaacaaaaatgtgaaacta 2149
QY 629 aaggcccttgccaggtgcttcagggtgctccgctgcttcttggtattaccggagcctgcgac 688
Db 2150 aaagggttgctgttcttcaagattaccgactatttttgaattactgaaccaactgta 2209
QY 689 ttccggtgtgaaccttcgctgcgctggccgttcttcttcacgtatcgtgacgcagctatc 748
Db 2210 tatggcgtgactttaccattgaaaaaacatttatttcagctgtgttgggtggcggtatc 2269
QY 749 ggtggcgttttgattgcaactcttataatcaaggcaactgctggtggcgctgcaggtttc 808
Db 2270 ggtgggtgcatgttggtgctatgaatcacgtgaaaaaacttaccgttggctgggttagtatg 2329
QY 809 ttgggtgtgtttctctattgatgctccagatatggtcatgttcttgggtgtgcaagtgttt 868
Db 2330 ttgagcttgctggctttattctcgcagagacaaaagatactcacogattgattactggt 2389
QY 869 accttcttcacgcattcgcgcgagcgttcttattgcttacttacttgggtgcgcgcaac 928
Db 2390 gcaattggtgcggaaattgctttatcatgctgttcttcttaacgtttgttcttactgtt-- 2447
QY 929 ggcagcattgattccagatgcaacgcgtgctccagtgctcgcgtcagagaacgcaacgcaaa 988
Db 2448 -----ttgaagtaacactatccagaacagcagactgaaaaaacagaaactgataag 2500
QY 989 gcagaagcaccgcagaaattttcaaacgattccacatcatccaggca---cctttgacc 1045
Db 2501 atggtggcaaccttaaaacgaaatcaagaagacaaaattatttagcaagtcacacttcaa 2560
QY 1046 ggtgaagctattgcactgagcagcgtcagcgtgccaatggtttccagcggaagcgtggc 1105
Db 2561 ggtgaattttaccgctagaaaaagtaacaagaccctgttttcttcagggtgcttttagga 2620
QY 1106 tcggcggttgccatgctcccaacaaaggcgagtcagttcttcctcgggtgagtgaaagatt 1165
Db 2621 aaagggtgtgcaattgagcgcagctgaaggcaaacctgtatgcaccgcagatggtgaaatc 2680
QY 1166 gtggtggcattccattcgtccatgcttccagctgctccaccagcagcgtgaggtgttcc 1225
Db 2681 accaattattccagacaggaacatgctgttggtgttgacgacaaacagagg----- 2729
QY 1226 aatgtggatatcttgatgcacattggtttcgcacacagtaaacctcaacgcgcagcacttt 1285
Db 2730 -gcgtgaaattattaatgcatattgcatgtgacggtcgaattagatggttaaaggcttt 2788
QY 1286 aaccgcgtgaagaagcaggcgtgaaatgaagtcaaaagcaggggagctgctgtgtgaattgat 1345
Db 2789 gaattatcagtgaaacaagggtgattctgttaaaaaaggagattgtctagttactttgat 2848
QY 1346 attgatgcattaaaggctgcaggttatgaggttaaccacgcgattgtgtttcgaattac 1405
Db 2849 attgctgcattaaaagaagcgtgttaccctcggtagttacaccgattgtgtgtaacgaatacg 2908
QY 1406 aa 1407
DB ||
DB 2909 aa 2910

```















Db 227 acagttattggtcccaaatgttattactatctaatgaatcacagatgctattgttgg 286  
QY 416 cttatgatttcggtggtccagtcggtgctgtcttctgttctgtctactcaacaatc 475  
Db 287 ttataatgaactgggtctcattaggaatggggatcttgggtgaacataattcaactcatt 346  
QY 476 gtcacactggtctcacaccagctcttcgcgcaattgactgagctgttt 526  
Db 347 gttatgactggtcttcacatcagtcattcttcgcatcgaaacacagtttactt 397

## RESULT 15

AAT47505

ID AAT47505 standard; DNA; 357 BP.

XX AC AAT47505;

XX DT 19-FEB-1998 (first entry)

XX DE Partial Enterobacter sp. scrA gene region.

XX KW scr operon; scrA; sucrose metabolising enzyme; sucrose isomerase;  
KW reduced metabolism; non-carbogenic sugar; trehalulose; palatinose;  
KW decomposition; monosaccharide; improved yield; disaccharide; ds.  
XX OS Enterobacter species.

XX PN EP751218-A2.

XX PD 02-JAN-1997.

XX PF 27-JUN-1996; 96EP-0110396.

XX PR 28-JUN-1995; 95DE-4023560.

XX PA (SUD-) SUEDEZUCKER AG MANNHEIM/OCHSENFURT.

XX PI Klein K, Mattes R, Stegmaier S;

XX DR WPI; 1997-054676/06.

XX PT Cell with reduced sucrose metabolism but with sucrose isomerase  
PT activity - providing high yields of non-carbogenic sugars, esp.  
PT trehalulose and palatinose

XX PS Claim 6; Page 14; 26pp; German.

XX CC The present sequence comprises a partial DNA sequence from the scr operon  
CC containing the scrA gene region. The scr operon encodes sucrose  
CC metabolising enzymes, in particular sucrose isomerase. Cells containing  
CC at least one DNA sequence encoding a protein with sucrose isomerase  
CC activity, and having reduced ability to metabolise sucrose, or their  
CC extracts, optionally in an immobilised form, are used to produce  
CC non-carbogenic sugars, especially trehalulose and/or palatinose. During  
CC production of the non-carbogenic sugars, decomposition of sucrose to  
CC monosaccharides (which are difficult to separate from the product) is  
CC minimised, so the yield of the disaccharides is improved.

XX SQ Sequence 357 BP; 46 A; 121 C; 107 G; 81 T; 2 other;

Query Match 4.7%; Score 71.6; DB 18; Length 357;  
Best Local Similarity 53.1%; Pred. No. 2.7e-11;  
Matches 152; Conservative 0; Mismatches 134; Indels 0; Gaps 0;

QY 183 tgtttggttttagatgttgcacagcggttaccaggccacggtcttctgtcgtggtgg 242  
Db 67 tcttcggcatcgaaagtggcgatgatcggtaccaggccacggtcttcccggtgctgctgg 126  
QY 243 ttcttggatttcgcaacgcatcgagaagtctctgcacagcgactcaagggcactgcag 302  
Db 127 cgggtggtttatgacatggtcgagaaacggtgcgcscgttatccctgacgcgctgg 186

QY 303 acttctgatactccagtgctgacgttctgctcaccggtattccttaccattcagcca 362  
Db 187 acctgatactcaccocgttctcgacggtgattatctccggtttatcgccctgctgctga 246  
QY 363 ttggccccagcaatgcgctgggtggcgatgctggcacacacggtctcacagggaactttatg 422  
Db 247 tcggcccgccggtcgcgctcgcgacggtatttcgtttatccctcagcagcgtttatca 306  
QY 423 atttcggtgttccagtcggcggtctgctgtctcttcgttctgtgcttactc 468  
Db 307 gccacgcccgtggtcggtggcggtgctgctgttcggcggtcttactc 352

Search completed: March 22, 2002, 08:53:35  
Job time: 4584 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: March 22, 2002, 06:41:31 ; Search time 2009.21 seconds  
(without alignments)  
12337.860 Million cell updates/sec

Title: US-09-604-231-1  
Perfect score: 1527  
Sequence: 1 ctctgcatctgcgcgtt.....gttgaaaccttgagttctgc 1527

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl:\*

- 1: gb\_ba:\*
- 2: gb\_htg:\*
- 3: gb\_in:\*
- 4: gb\_om:\*
- 5: gb\_ov:\*
- 6: gb\_pat:\*
- 7: gb\_ph:\*
- 8: gb\_pl:\*
- 9: gb\_pr:\*
- 10: gb\_ro:\*
- 11: gb\_sts:\*
- 12: gb\_sy:\*
- 13: gb\_un:\*
- 14: gb\_vi:\*
- 15: em\_ba:\*
- 16: em\_fun:\*
- 17: em\_hum:\*
- 18: em\_in:\*
- 19: em\_om:\*
- 20: em\_or:\*
- 21: em\_ov:\*
- 22: em\_pat:\*
- 23: em\_ph:\*
- 24: em\_pl:\*
- 25: em\_ro:\*
- 26: em\_sts:\*
- 27: em\_sy:\*
- 28: em\_un:\*
- 29: em\_vi:\*
- 30: em\_htgo\_hum:\*
- 31: em\_htgo\_inv:\*
- 32: em\_htgo\_rod:\*
- 33: em\_htg\_hum:\*
- 34: em\_htg\_inv:\*
- 35: em\_htg\_rod:\*
- 36: em\_htg\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1527	100.0	1527	6	AX069134	AX069134 Sequence
2	1515	99.2	349980	6	AX127152	AX127152 Sequence
3	1492	97.7	1983	6	AX122988	AX122988 Sequence
4	1107.4	72.5	1109	6	AX069136	AX069136 Sequence
5	287.6	18.8	21838	1	PDGRAFOPER	L32093 Pedicoccus
6	287.6	18.8	21839	1	PDSURPOP	L32093 Pedicoccus
7	250.8	16.4	5800	1	LL297015	Z97015 Lactococcus
8	231.8	15.2	10264	1	AE004395	AE004395 Vibrio ch
9	230.8	15.1	2508	1	STRSCRA	M2711 Streptococ
10	218.6	14.3	10085	1	AE006222	M76768 Vibrio algi
11	215.2	14.1	2955	1	VIBSCRAK	X69800 S.xylosus s
12	210.6	13.8	2655	1	SXSCRA	Z54245 B.subtilis
13	205.4	13.5	4158	1	BSTREAPR	D83967 Bacillus su
14	201.8	13.2	22197	1	D83967	D86417 Bacillus su
15	201.8	13.2	37900	1	D86417	Z99108 Bacillus su
16	201.8	13.2	208430	1	BSUB0005	AF229829 Pseudomon
17	194.8	12.8	341350	1	AP003365	AP001513 Bacillus
18	190.6	12.5	2913	1	AF269422	AP001508 Bacillus
19	190.6	12.5	2913	6	AX144742	X57401 K.pneumonia
20	183	12.0	299850	1	AP001514	AJ250722 Erwinia a
21	157.6	10.3	4487	1	AF229829	M81772 Erwinia chr
22	146.8	9.6	292550	1	AP001513	U34876 Bacillus st
23	146.4	9.6	296950	1	AP001508	D37921 Alkalophilol
24	143.6	9.4	4621	1	KPSCRYAB	X67750 S.typhimuri
25	141.8	9.3	6950	1	EM250722	J03006 B.subtilis
26	138	9.0	5065	1	ERWBGPA	X73124 B.subtilis
27	136.2	8.9	1398	1	BSU34876	Z99123 Bacillus su
28	135.2	8.9	2819	1	BACISPO	Y00541 Salmonella
29	134.6	8.8	4322	1	STSCRCOMP	AL449924 Streptoco
30	132.2	8.7	3518	1	BACSACP	AE007479 Streptoco
31	132.2	8.7	97015	1	BSENR	AF206272 Streptoco
32	132.2	8.7	212150	1	BSUB0020	AF205034 Clostridi
33	130	8.5	1530	1	STSCRA	AE007557 Clostridi
34	126.6	8.3	14515	1	AE007479	AF206272 Streptoco
35	123.4	8.1	151947	2	SPNEU1902	AL449924 Streptoco
36	120.8	7.9	8006	1	AF206272	AF206272 Streptoco
37	117.8	7.7	5600	1	AF205034	AF205034 Clostridi
38	117.8	7.7	10811	1	AE007557	AE007557 Clostridi
39	114.2	7.5	136254	1	ECOUW82	LI0328 E. coli; th
40	112.8	7.4	10371	1	AE007465	AE007465 Streptoco
41	111.8	7.3	2134	1	ECOBGLC	M15746 E.coli bglC
42	111.8	7.3	5270	1	ECOBGLC	M16487 E.coli bgl
43	111.8	7.3	12989	1	AE000449	AE000449 Escherich
44	111.2	7.3	300750	1	AP001509	AP001509 Bacillus
45	109.6	7.2	3615	6	AR012068	AR012068 Sequence

ALIGNMENTS

RESULT 1	AX069134	AX069134	1527 bp	DNA	PAT	25-JAN-2001
LOCUS	Sequence	1 from Patent	WO0102583.			
DEFINITION	Sequence	1 from Patent	WO0102583.			
ACCESSION	AX069134					
VERSION	AX069134.1	GI:12579016				
KEYWORDS						
SOURCE						
ORGANISM						

Corynebacterium glutamicum.  
Corynebacterium glutamicum  
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
Actinomycetales; Corynebacteriineae; Corynebacteriaceae;  
Corynebacterium.

REFERENCE 1 (bases 1 to 1527)  
AUTHORS Pompejus, M., Kroeger, B., Schroeder, H., Zelder, O. and Haberhauer, G.  
TITLE Corynebacterium glutamicum genes encoding phosphoenolpyruvat e:  
sugar phosphotransferase system proteins  
JOURNAL Patent: WO 0102583-A 1 11-JAN-2001;  
BASf AKTIENGESSELLSCHAFT (DE)  
FEATURES Location/Qualifiers  
SOURCE 1. 1527

/organism="Corynebacterium glutamicum"

/db\_xref="taxon:1718"

101. 1507

/note="RXS00315"

/codon\_start=1

/transl\_table=11

/protein\_id="CAC27276.1"

/db\_xref="GI:12579017"

/translation="MAWVPSLVNGYDAATMAAGEMPWLSFLGLDVAAGVGTVLP  
VLVYSILATIEFLHKLKGTADFLITPVLTLLLLGFLTFIAIGPAMRWGDLVLAHG  
LQGLYDFGPGVGLLGLVSPVITGLHQSFPPIELFLNFGSGFIATASWANIAG  
GAALFAFLKALSEKILAGASGVSAVLGTEPAIFGVNLRWPFPIGIGTAIIGG  
ALITALFNIKAVLAAGFLGVSIDAPVMFLVCAVFFIAFAAAIAYGLVLRN  
GSDIDPATVPAGVTKAEAEAPAEFSNDSTIIQAPLTGEALASVSDMFPASCKL  
GSGVATPPTKGLVSPVSGKIVVAFPSGHAFVTRKAEDEGSNDVILMLHIGDFVNLNG  
THPNLKKQDEKAGELICEFDIDAIRKANGYEVVTPVIVSNVKKTPVNTYGLGEIE  
AGANLLNVAKEAVPAIP"

BASE COUNT 304 a 392 c 430 g 401 t

ORIGIN

Query Match 100.0%; Score 1527; DB 6; Length 1527;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1527; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	ctcatggcatctgcgcgcttcgqcttcctgccagatgttggtggtttccaccgcaaccaag	60
DB	1	CTCATGCGATCTGCGCGCTTCGCGTCTTCCAGTGTGGTGGTTCACCGCAACCAAG	60
QY	61	cgttctggcggaatagttctctggcgccgctgattggtggtttccaccgcaaccaag	120
DB	61	CGTTTCGCGGCAATAGATTCTCGGCGCGCGTATTGGTATGCGGATGGTGTTCGCGAG	120
QY	121	cttggtgaacgctacacatggtccgccaccatgctgcggcgcaaatgccatgtgctc	180
DB	121	CTTGGTGAACGGCTACACATGTCGCGCGCCACCATGGCTGCGCGGCAATGCCAATGTGCTC	180
QY	181	cctgtttggttttagatgttcccaagccggttacagggacacgctgcttctctgctggt	240
DB	181	CTGTGTTGGTTTAGATCTTGCCCAAGCCGTTACACAGGCAACCGTCTCTGTGCTGGT	240
QY	241	ggtttcttgattctggaacatcgagaggttctctgcacacgactcaaggcactgc	300
DB	241	GGTTTCGTTGGATTCTGCAACGATCGAGAAGTTCTCTCACAAAGCGACTCAAGGGCACTGC	300
QY	301	agacttctactactcactcagctgcagcttgcctcactcaccgactccttacattcctgc	360
DB	301	AGACTTCTGATCACTCCAGTGTGACGTTGCTGCTCACCAGGATTCCTTACATTTCATCGC	360
QY	361	cattggcccagcaatgcgctgggtggcgatgtgtgcacacggtctacagggaacttta	420
DB	361	CATTGGCCCAAGCAATGCGTGGGTGGCGATGTGCTGGCACACGGTCTACAGGGACTTTA	420
QY	421	tgatttcggttggtccagtcggcggtctgctctcgttcggttcggttactcaccacatcgcat	480
DB	421	TGATTTCCGTTGGTCCAGTCCGCGGCTGCTCTTCGGTCTGCTACFCACCAATCGTCAT	480
QY	481	cactggtctcacagtcctctccgcacaaattgagctgagctgtttaaccagggtggtatc	540
DB	481	CACTGGTCTGCACCACTCTTCCCGCAATTGAGCTGGAGCTGTTTAACCAAGGTTGGATC	540
QY	541	cttcattctgcgaacgcatctatggctaataatcgccacgggtgcgcatgtttggcagt	600
DB	541	CTTCATCTTCGAACGGCATCTATGGCTAATATCGCCACGGGTGCGGCATGTTTGGCAGT	600
QY	601	gttcttctggcgaagaagtgaagctcaaggcccttgaggctgcttcacaggtgtctccgc	660
DB	601	GTTCTTCTCGCGAAGAGTGAAGAGCTCAAGGGCTTGCAGGCTGCTTCAGGTGCTCCCG	660
QY	661	tgttcttggttatcagcagctcgatcttcggtgtgaaaccttcgctgcgctggcggtt	720
DB	661	TGTTCTTGTTATTACGAGGCTCGGATCTTCGGTGTGGAACCTTCGCGCTGCGCTGGCGGTT	720

QY	721	cttcacggtatcggtaccgcagctacggtggtggtggttcttattgacactctttaaatacaa	780
DB	721	CTTCATCGGTATCGGTACCGCAGCTATCGGTGGCGCTTTGATTCGACTCTTTAATATCAA	780
QY	781	ggcagttgcttggtggcgctgcaggtttcttgggtgtgttcttattgatgtccagatat	840
DB	781	GGCAGTTGCGTGGCGCTGCAGGTTTCTTGGGTGTGTCTTATTGATGCTCCAGATAT	840
QY	841	ggtcatgttttgggtgtgtgagttgttaactcttcttcacgtcgcgcagcagattgc	900
DB	841	GGTCATGTTTCTTGGTGTGTGAGTTGTACTCTTTCATCGCATTCGGCGCAGGATTC	900
QY	901	ttatggccttacttgggttcgcgcgaacgagcagcattgatccagatcaaacgctgtcc	960
DB	901	TTATGGCCTTACTTGGTTCGCCGACGCGCAGCATTTGATCCAGATGCAACCGCTGTCC	960
QY	961	agtctctgcgggaacgacccaagccgagcagaagcaccgcgagaatttttcaaacgattc	1020
DB	961	AGTGCCTGTCAGAACGACCAAGCCGAGAGCAGACCCGCGAGATTTTCAACGATTC	1020
QY	1021	caccatcatcagcaccctttgaccggtgaagctatttgacacgaagcagcgtcagcgatgc	1080
DB	1021	CACCATCATCCAGCACCTTTGACCGGTGAAGCTATTGCACTGAGCAGCGTCACCGATGC	1080
QY	1081	catgtttgcagcggaagcttgctcggtggttgcctatcgctcccaaccgaaggcgagtt	1140
DB	1081	CATGTTTGCAGCGGAAGCTTGGCTCGGCGTTCGCCATCGTCCCAACCAAGGGCGAGTT	1140
QY	1141	agttctccggtgagtggaagattgtgtggtggaattcccatctggccatgcttttcgagtt	1200
DB	1141	AGTTTCTCCCGTGAGTGAAGATTTGTTGGCATTTCCCATCTGGCCATCTTTCGCGAGT	1200
QY	1201	tcgcaccaagctgagtgatggttccaatgtggatatcttgatgcacattggtttcgacac	1260
DB	1201	TCGCACCAAGCTGAGGATGTTCCCAATGTGGATATCTTGTATGCACATTTGGTTTCGACAC	1260
QY	1261	agtaacctcaacgagcagcactttaaccgctgaaagaacgagcgatgaagtcgaagc	1320
DB	1261	AGTAACCTCAACGGCACGCGACTTTAAACCGCTGAAGAAGCAGGGCGATGAAGTCAAAGC	1320
QY	1321	agggagctgctgtgaaatcgatatgtgatgccattagcattagggctgaggttatgaggtaac	1380
DB	1321	AGGGAGCTGCTGTGTGAATTTCGATATTGATGCAATTAAGGCTGCAGGTTATGAGGTAAC	1380
QY	1381	cagcgcatgtgttctcgaattacaagaaacccgacctgtaaacacttacggtttggg	1440
DB	1381	CAGCGCATTTGTTGTTTCGAATTACAAGAAACCCGGACCTGTAAACACTTACGGTTTGGG	1440
QY	1441	cgaaatgaagcgggagccaacctgctcaacgtcgcaaaagaaagcagcgtgcccagcaac	1500
DB	1441	CGAAATGAAGCGGGAGSCCAACCTGCTCAACGTCGCAAGAAAGAAAGCGGTGCCAGCAAC	1500
QY	1501	accataagttgaacaccttgagttgctg 1527	
DB	1501	ACCATAGTTGAACCTTGAGTGTTCG 1527	

RESULT 2

AX127152/c

LOCUS

AX127152 349980 bp DNA

Sequence 7068 from Patent EP1108790.

AX127152 AX114121

AX127152.1 GI:14041140

VERSION

KEYWORDS

SOURCE

ORGANISM

Corynebacterium glutamicum.

Corynebacterium glutamicum

Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

Actinomycetales; Corynebacterineae; Corynebacteriaceae;

Corynebacterium.

1 (bases 1 to 349980)

Nakagawa, S., Mizoguchi, H., Ando, S., Hayashi, M., Ochiai, K.,

Yokoi, H., Tateishi, N., Senoh, A., Ikeda, M. and Ozaki, A.

Novel polynucleotides

TITLE

PAT 11-MAY-2001

JOURNAL	Patent: EP 1108790-A 7068 20-JUN-2001;
FEATURES	KYOWA HAKKO KOGYO CO., LTD. (JP)
source	Location/Qualifiers 1. .349980 /organism="Corynebacterium glutamicum" /db_xref="taxon:1718" /note="Seq 1 to long (3.309.400) split in 11, seq 7068 2.700.001 3.049.980"
BASE COUNT	81250 a 97718 c 90621 g 80391 t
ORIGIN	
Query Match	99.2%; Score 1515; DB 6; Length 349980;
Best Local Similarity	99.9%; Pred. No. 0;
Matches 1526; Conservative	0; Mismatches
Qy 1	cctatggcatctgcgcccgttcgtttcgcagtggttgcacccaacaag 60
Db 111326	CTCATGGCCATCTGCGCGTTTGGGTTCCTTGCCAGTGTGGTGGTTACCCGAACAAG 111267
Qy 61	cgtttccggcgcaatgagtcttcctggcgccgcgatattggtatggcgatggtgtcccgag 120
Db 111266	CGTTTCGGCGCAATGAGTTCCTTGGGCGCG -GTAATTGGTATGGCGATGTTCCCGAG 111208
Qy 121	cttgggtgaacggtactagaogtggccgccaccatacctggctgcggcgaaatgccaatgtggtc 180
Db 111207	CTTTGGTGAACGGTATGACATGGTGCCTGCCACCACATGGCTGGCGGAAATGCCAATGTGGTC 111148
Qy 181	cctgtttggttagatgtgtgccaacgcggtttaccaggcacccggtcttcctgtgctggt 240
Db 111147	CCGTGTTGGTGTAGATGTGGCCCAAGCCGGTTACCAAGGCACCGTGTTCCTGTGCTGGT 111088
Qy 241	ggttcttgtgattctggcaacgacatcgagaagttccctgcacaaagcagctcaagggaactgc 300
Db 111087	GGTTCTTGGATTCTGGCACAGATCAGAAGTTCCTTGCAACAGCACTCAAGGGCACTGC 111028
Qy 301	agacctcctgatacactecagtcgtaacgttgctgcgtccacggattccttacattcatcgc 360
Db 111027	AGACTTCTGATCCTCAAGTGTGCTGCTGCTCAGCGGATTCCTTACATTTCATCGC 110968
Qy 361	cattggccgaagcaatgcgttgggtggcgatgtgtgcacacaggtctcacagggaacttta 420
Db 110967	CATTGGCCCCAGCAATGCGGTGGGTGGCGATGTGCTGGCACAGGTCCTACAGGGACTTTA 110908
Qy 421	tgattcgggtgggtccagtcggcggtctgctctcggctgctggtctactcaccaatcgtcat 480
Db 110907	TGATTTCGGTGGTCCAGTCGGCGGTCTGCTCTTCGGTCTGGTCTACTCACCATCGTCAT 110848
Qy 481	cactggtctgcacagtccttccccccaatttgagctggagctgttttaaccagggtggatc 540
Db 110847	CACTGGTCTGCACAGTCCCTCCGCCAATTGAGCTGGAGCTGTTTAAACCAGGGTGGATC 110788
Qy 541	cttccattctgcgaacggcatctatgctaataatcgcgccagggtgcggcatgttttggcagt 600
Db 110787	CTTCATCTTCGC AACGGCATCTATGCTATAATATCGCCAGGTCGGCGCATGTTTGGCAGT 110728
Qy 601	gttctctcggcaagagtgtaaagctcaaggcccttcagggtgcttcagggtgtctccgc 660
Db 110727	GTTCTTCTGGCGAAGAGTGAAAGCTCAAGGGCCTTGACAGTGTCTCAGGTGCTCCGC 110668
Qy 661	tgttcttggtattcacggagcctgcgatcttctggtgtgaaaccttcgctgcgctggcgttt 720
Db 110667	TGTTCTTGGTATTACGAGCGCTCGCATCTTCGGTGTGAACCTTCGCTCGCTGGCGCGTT 110608
Qy 721	cttccatcggtatcggtaccgcagctatcgggtggcgctttgattggaactotttaatatcaa 780
Db 110607	CTTCATCGGTATCGGTACCCGACCTATCGTGGCGCTTTGATTGCACTCTTTAATAATCAA 110548
Qy 781	ggcagttgcttgggcgctcaggttcttgggtgtgtttcttatgatacgtcccaagatat 840
Db 110547	GGCAGTTCGGTGGCGCTCCAGGTTCTTGGGTGTTGTTTTATTGATGCTCCAGATAT 110488
Qy 841	ggtcagtattcgtgtgtagcagtttgttaccttcttcacgttcgacttcggcgagcgaggttc 900

Db 110487	GGTCATGTTCTTGGTGTGTGCAGTTGTTACCTTCTTCATGCATTCGGCGCAGGATTCG 110428
Qy 901	ttatggcctttacttggttgcgcgcaacgcgcagcatigatccagatgcaaccgctgtctcc 960
Db 110427	TTATGGCCCTTTACTTGGTTCGCCGCAACGGCAGCATTTGATCCAGATCAACCGCTGCTCC 110368
Qy 961	agtgcctgcagaaacagcaccacaaagccgaagcagaagacccgcagaaattttccaacgattc 1020
Db 110367	AGTGCCTGCAGAAACGACCAAGCGAAGCAGAAAGCAGCCGACAGATTTTCAAACGATTC 110308
Qy 1021	caccatcatcaggcacccctttgaccggtgaagctattgcactgagcagcgtcagcgatgc 1080
Db 110307	CACCATCATCCAGGCACOCCTTTGACCGGTGAAGCTATTGCATGAGCAGCGTCAGCGATGC 110248
Qy 1081	catgtttccgcgcgaaaaagcttggctgcggcggttggccatgctcccaaccgaaggcgagtt 1140
Db 110247	CATGTTTCCGACGGGAAGCTTGGCTCGGGCGTTTGCCATCGTCCCAACCAAGGGGCAATT 110188
Qy 1141	agtttctccggtgagtgaaaagattgtgtggcattcccatctggccatgctttcgcagt 1200
Db 110187	AGTTTCTCCGCTGAGTGGAAAAGATTGTGTGGCATTTCCCATCTGGCCATGCTTTTCGCAGT 110128
Qy 1201	tcgcaccaagcctagagatggttcccaatgtggatatatttgatgcacatttggtttcgcacac 1260
Db 110127	TCGCACCAAGCGCTGAGGATGTTCCAATGTGGATATCTTGTATGCACATTTGTTTCGACAC 110068
Qy 1261	agtaaacctcaacgcgcgcacactttaaccgcctgaagaagcagggcgatgaagtcaaaagc 1320
Db 110067	AGTAAACCTCAACGCACGCACACTTTAACCCGCTGAAGAANGCAGGGCGATGAAGTCAAAAGC 110008
Qy 1321	aggggagctgctgtgtaattcgatattgatgccattaaagctcagaggttatgaggtlaac 1380
Db 110007	AGGGGAGCTGCTGTGTGAATTTCGATATTGATGCCATTAAAGCTTCAGGTTATGAGGTAAC 109948
Qy 1381	cacgcgcgattgtgttttcgaattacaagaaacccggacccctgtgaacacttaacacttgcgttgg 1440
Db 109947	CACGCGGATTTGTTTTCGAATTACAGAAAAACCGGACCTCTAAACACTTTACGGTTTTGGG 109888
Qy 1441	cgaattgaagcggagcgaacacctgcctcaacgtcgcaaaagaaagcgggtgccagcaac 1500
Db 109887	CGAAATTGAAGCGGAGCCACCTGCTCAAGCTCGAAAGAAAGAGCGGTGCCAGCAAC 109828
Qy 1501	accataagttgaaaccttgagttgctg 1527
Db 109827	ACCATAAGTTGAACCTTGAGTGTTCG 109801
RESULT 3	
AX122988	
LOCUS	AX122988 1983 bp DNA PAT 11-MAY-2001
DEFINITION	Sequence 2904 from Patent EP1108790.
ACCESSION	AX122988
VERSION	AX122988.1 GI:14040476
KEYWORDS	Corynebacterium glutamicum.
SOURCE	Corynebacterium glutamicum.
ORGANISM	Corynebacterium glutamicum
REFERENCE	1 (bases 1 to 1983)
AUTHORS	Nakagawa,S., Mizoguchi,H., Ando,S., Hayashi,M., Ochiai,K., Yokoi,H., Tateishi,N., Senoh,A., Ikeda,M. and Ozaki,A.
TITLE	Novel polynucleotides
JOURNAL	Patent: EP 1108790-A 2904 20-JUN-2001;
FEATURES	KYOWA HAKKO KOGYO CO., LTD. (JP) Location/Qualifiers 1. .1983 /organism="Corynebacterium glutamicum" /db_xref="taxon:1718"
BASE COUNT	410 a 508 c 558 g 507 t
ORIGIN	

[illegible]

QY	1021	caccatcatcaggccacotttgaccgggtgaagctattgcaactgaacgctgaagcgtgc	1080
Db	1500	CACCAATCATCCAGGCACACCTTTGACCGGTGAAGCTATTGCACTGACGACGCTCAGCGATGC	1559
QY	1081	catgtttgccacgcgaagacttgctcgggcggtgaccatcgctcccaaccgaaggcgcaatt	1140
Db	1560	CATGTTTGCACGGGAAGCTTGGCTCGGGCGTTGCCATCGTCCCAACCAAGGGCAGTT	1619
QY	1141	agttcttcgggtgagtggaaagattggtggcattcccatctggccatgctttcgcagt	1200
Db	1620	AGTTTCTCCGGTAGTGGAAAGATTGTTGTTGGCATTTCCCATCTGGCCATGCTTTGCGCAGT	1679
QY	1201	tgccaccaagcgtgaggttggtccaatgtggatatcttgatgcacattggtttcgacac	1260
Db	1680	TCGCACCAAGGCTGAGGATGGTTCCAAATGTGGATATCTTGATGACATTTGTTTTCGACAC	1739
QY	1261	agtaaacctcaacgcacgcacgtttaaccctgaagcctgaagaagcaggcgcatgaagtcaaaagc	1320
Db	1740	AGTAAACTCTAACGGCAGCGACTTTTAAACCCGCTGAAGAAGCAGGCGGATGAAGTCAAAGC	1799
QY	1321	aggggagcgtgctgtgtaattcgattgatgccaattgaaggctgcaggtcttatgaggtaac	1380
Db	1800	AGGGAGAGTGCTGTGTAATTCGATATTATGTCSCATTAAAGGCTGCAGGTTATGAGGTAAC	1859
QY	1381	cacgcgcgattgttttcgaattacaagaacacggacgactgaacacttaacggtttggg	1440
Db	1860	CACGCGCATTTGTTTTCGAATTACAAGAACAACCGGACCTGTAAACACTTACGGTTTGGG	1919
QY	1441	cgaattgaacgggagcgaacctgctcaacgtgcgaagaagaagcgggtgcacgaac	1500
Db	1920	CGAAATTGAAGCGGGAGCGCAACCTGCTCAACGTCGCAAGAAGAAGCGGTGCCAGCAAC	1979
QY	1501	acca 1504	
Db	1980	ACCA 1983	
RESULT	4		
AX069136			
LOCUS	AX069136	1109 bp	DNA
DEFINITION	Sequence 3 from Patent WO0102583.	PAT	25-JAN-2001
ACCESSION	AX069136		
VERSION	AX069136.1	GI:12579018	
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS	Pompejus,M., Kroeger,B., Schroeder,H., Zeider,O. and Haberhauer,G.		
TITLE	Orynebacterium glutamicum genes encoding phosphoenolpyruvat e:		
JOURNAL	sugar phosphotransferase system proteins		
Patent:	WO 0102583-A 3 11-JAN-2001;		
FEATURES			
source	BASF AKTIENGESellschaft (DE)		
	Location/Qualifiers		
	1..1109		
	/organism="Corynebacterium glutamicum"		
	/db_xref="taxon:1718"		
CDS	<1..1089		
	/codon_start=1		
	/transl_table=11		
	/protein_id="CAC27277.1"		
	/db_xref="GI:12579019"		
	/translation="YDFGVPVGGLLFGLVSPVITGLHQSPFPIELELFNQGGSFIF		
	ATASMANIAQCAACLAFFVILAKSEKLGAGSGVSAVLGITGPATFGVNLRLWRPF		
	ITGTTAAIGGALLTALFIRKAVALLGAGFLGWSIDAPDMVMLFVCAVVTFFIAFGAAI		
	AYGLYLVRNGSDIDPATAAPVPAGTTKAEAPAEFSNDSTLIQAPLTGEATLSLV		
	SDAMFASCKLGSGLVAIVPTKGQLVSPVSGKIVAFVPSCHAFVARTKADGDSNVIDLHMH		
	IGDFTVNLNGTHENPLUKQGVDEKAGELLCEFDIDAIKAAGYEVTTPIVVSNYKKTGP		
	VNTYGLGEIEAGANLNLNVAKEAVPATP"		
BASE COUNT	236 a	278 c	292 t



ORIGIN

Query Watch 72.5%; Score 1107.4; DB 6; Length 1109;  
Best Local Similarity 99.9%; Pred. No. 3.7e-289;  
Matches 1108; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 419 tatgatttcgggtggtccagtcggtgctgctcttcttggtctgctactcaacaatcgtc 478  
|||||  
Db 1 TATGATTTGGCGGTGACGTCGCGGTGCTCTCTCGGTGCTGCTACTCAACAATCGTC 60  
  
QY 479 atcactgggtctgcaccagtccttccgcgaattgactgagctgtttaaccagggtgga 538  
|||||  
Db 61 ATCACYGGTCTGCACCAAGTCTTCCCGCCCAATGACCTGGAGCTGTTTAACCAAGGTGGA 120  
  
QY 539 tccttcattctgcacagcgcatctatggctaa tatcgccaggggtcggcgatgtttggca 598  
|||||  
Db 121 TCCTTCACTCTCGCAACGGCATCTATGGCTAAATATGCCCAGGGTGGCATGTTGGCA 180  
  
QY 599 gtatttctctggcgaagagtgaaagctcaaggccttgcaaggtccttcagggtgctcc 658  
|||||  
Db 181 GTGTTCTTCTGCGGAAGAGTGAAGAGCTCAAGGGCTTTCGAGTGCTTCAGTGCTCC 240  
  
QY 659 gctgtcttggattacgagagcctgcgatcttcggtgtaaccttcgctgcactgcgcg 718  
|||||  
Db 241 GCTGTTCTTGGTATTACGGAGCTCGCATCTTCGGTGTGAACCTTCGCCCTGCGCTGCCG 300  
  
QY 719 ttcttcacgtgatcggttaaccgcagctatcggtggcgcttggattgacactctttaatc 778  
|||||  
Db 301 TTCTTCACTCGGTATCGGTACCGCAGCATATCGGTGGCGCTTGTATTGCACCTTTAATATC 360  
  
QY 779 aaggcagctgctgtggcgctgcaggtttcttgggtgtgtttcttattgatgtccagat 838  
|||||  
Db 361 AAGCGAGTTCGTTGGCGCTGCAGGTTCCTGGGTGTTGTTCTATTGATGCTCCAGAT 420  
  
QY 839 atggtcatgtcttgggtgtgagttgttaccttcttcacgtcattcgccgagcgatt 898  
|||||  
Db 421 ATGTCATGTTCTTGGTGTGTCAGTTGTACCTTCTCATCCATTCGCGCAGCGATT 480  
  
QY 899 gctatggccttacttggttcgcgcaacgagcagcttgatgacatgcaacgcgtgct 958  
|||||  
Db 481 GCATTATGGCTTTTACTTGGTTCGCCCGCAGCGCAGCATTCAGATGCAACCGCTGCT 540  
  
QY 959 ccagtcctcagaaacgaaacgaaagcagaagcagcccgacagatttccaacgat 1018  
|||||  
Db 541 CCAGTGCCTCAGAAACGACCAAGCCGGAAGAGCAGACGCCGCAATTTTCAACAGAT 600  
  
QY 1019 tccaccatcatccaggcacctttgaccggtgaagctattgactgagcagcgtcagcgat 1078  
|||||  
Db 601 TCCACCATCATCCAGGCACCTTTGACCGGTGAAGCTATTGCACTGAGCAGCGTCAGCGAT 660  
  
QY 1079 gccatgtttgcgagcgaagcttgctcggcgttgccatctcccaacccaaggggcag 1138  
|||||  
Db 661 GCCATGTTGCCACGCGAAAGCTTGGCTCGGGGTTGCCATCTGCCCATGCTTCGCA 720  
  
QY 1139 ttagtcttcctgggtgagtgaaagattgtggtgcatctccactcctggccatgtttcgca 1198  
|||||  
Db 721 TTAGTTTCTCCGGTGAGTGAAGAGATTGGTGGCATTCCTCATCTGCCCATGCTTCGCA 780  
  
QY 1199 gttcgcaccaaggtcaggatgttccaatgtgatatcttggatcttgatgacattgtttcgac 1258  
|||||  
Db 781 GTTCCGACCAAGGCTGAGGATGTTCCAAATGTGATATCTTGTATGCACATGTTTCGAC 840  
  
QY 1259 acagtaaacctcaacggcagcactttaaccgcctgaagaagcaggcgatgaagtcaaa 1318  
|||||  
Db 841 ACAGTAAACCTCAACGCGCAGCCTTTAAACCCGCTGAAGAAGCAGCGCATGAAGTCAA 900  
  
QY 1319 qcagggagcgtcgtgtgaattcgatattgatgccattaaagctcaggtttatgaggtta 1378  
|||||  
Db 901 GCAGGGAGCTGCTGTGTAATTCGATATTGATGCCATTAAAGGCTGCAGGTTATGAGGTA 960  
  
QY 1379 accacgcgattgttttcgaattacaagaacacggacgtgtaaacacttacggtttg 1438  
|||||

Db 961 ACCACGCCGATTGTTGTTTCGAATTACAAGAAACCCGACCTCTAAACACTTACGGTTTG 1020  
  
QY 1439 ggcgaattgaagcgggagccaaacctgctcaacgctgcgaagaagaagcgggtgccagca 1498  
|||||  
Db 1021 GCGCAATTGAAGCGGAGCAACCTGCTCAGCTGCCAAAGAAAGCGGTGCCAGCA 1080  
  
QY 1499 acacataagttaaaaccttgagttgctg 1527  
|||||  
Db 1081 ACACATAAGTTGAAACCTTGAAGTGTTCG 1109

RESULT 5  
PCR/AFOP/ER/c  
LOCUS PCR/AFOP/ER 21838 bp DNA BCT 23-MAY-1994  
DEFINITION Pediococcus pentosaceus raffinose operon genes.  
ACCESSION L32093  
VERSION L32093.1 GI:493181  
KEYWORDS agar gene; ags gene; agl gene; alpha-galactosidase;  
alpha-glucosidase; fructokinase; insertion element; permease; rafp  
gene; rafr gene; raffinose operon; regulatory protein; scrA gene;  
scrB gene; scrR gene; scrR gene; sucrose-6-phosphate;  
transport protein.  
SOURCE Pediococcus pentosaceus (strain PPE1.0) DNA; Insertion sequence  
IS30 homolog (transposable element Insertion sequence IS30 homolog,  
kingdom Prokaryotae) DNA; Pediococcus pentosaceus (strain PPE1.0)  
DNA; Insertion sequence IS3 homolog (transposable element Insertion  
sequence IS3 homolog, kingdom Prokaryotae) DNA; and Pediococcus  
pentosaceus (strain PPE1.0) DNA.  
ORGANISM Pediococcus pentosaceus  
Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;  
Pediococcus.  
REFERENCE 1 (bases 1 to 21838)  
AUTHORS Leenhouts, K.J., Bolhuis, A.A., Kok, J.J. and Venema, G.G.  
TITLE The sucrose and raffinose operons of *Pediococcus pentosaceus* PPE1.0  
JOURNAL Unpublished  
COMMENT On May 25, 1994 this sequence version replaced gi:475106.  
FEATURES  
Location/Qualifiers  
1..21838  
/organism="Pediococcus pentosaceus"  
/strain="PPE1.0"  
/db\_xref="taxon:1255"  
487..510  
/note="DR2; putative"  
/rpt\_type=direct  
488..561  
/note="IR1; putative"  
/rpt\_type=direct  
513..561  
/note="DR1; putative"  
/rpt\_type=direct  
2724..2760  
/standard\_name="terminator 4"  
/note="putative"  
complement(2929..3864)  
/gene="rafr"  
complement(2929..3762)  
/gene="rafr"  
/function="regulation raffinose operon"  
/note="putative"  
/codon\_start=1  
/transl\_table=11  
/product="regulatory protein"  
/protein\_id="AA025562.1"  
/db\_xref="GI:475107"  
complement(3837..3842)  
/gene="rafr"  
complement(3859..3864)  
-10\_signal  
-35\_signal

gene	/gene="rafr" 3909..5918 /gene="rafp" 3909..3914 /gene="rafp" /note="putative" 3932..3937 /gene="rafp" 3993..5918 /gene="rafp" /function="raffinose transport protein" /note="putative" /codon_start=1 /transl_table=11 /product="permease" /protein_id="AAA25563.1" /db_xref="GI:475108" /translation="MQEENHYKMGGRLYGFGCAKNDAPYSILSGYLIIFTTSHLFD TGNKALDNRMVSLVTLIIIMVLRIVELFIDPFIGNAIDRTKNSPGHFRPWVVGTVSS IITLLFTNLGLYAKNMIYLVVFALYITIMFYDFSKDVGFWMLPSLTTDSRERE KTAFTARLSTGGGLVGVMPVAFVFFSAKATSTGDNRGVTFALICIALISAWG VGLTYREVSJDIRKNKQDTPGVMEIFFKALAKNDOLLAAALVYFGVGINILGSLEVY XYTILMGKPKSILSIIINFLGLIATSLPEVLSKFKESFGAGCLVMSGLAIIFT IAGSNILVLLAATWFGFPQQMWFLVVLVITDSVEYGGKLGHREDSUALSVRPLID KFGGAISSNGVOAIIISGMTGTGATASSITAAQGLHFKTLTFMFAFAPMLLIATIFESK QIFTEBKEHAEIVALERTWTKFDNTGDAEKVYTSLDLATPIAGQVILPQAQVNDP TAAAGTLGGFAIKPSDGRILAPEDATVQVFTTRHAGVLGVNGIVLLIHHGLGVK LRGTGFISSVEGQHVQCGDELLEFWDPTIKQAGLDDTVIMTVTNSTEFTMMDBWLKP GQAVKATDNILQHLTKA" 5976..8177 /gene="agar" 5976..8177 /gene="agar" /codon_start=1 /transl_table=11 /evidence="experimental" /product="alpha-galactosidase" /protein_id="AAA25564.1" /db_xref="GI:475109" /translation="MSLIIVDQANRVFHLHNQTLSYIFAVEQGGTSLHLYFGGHVDHY HGELRTPVRDGFSGNLPTDFTSRDLPKPEYSTAGENDYHLPAIVRHTDGNAL YLVYQVQCNQGHAVNLEKPLSPAAVDETEAETLIVLPQVSQVEEDLQYTIYDRPV VTRSVQVQCNQGHAVNLEKPLSPAAVDETEAETLIVLPQVSQVEEDLQYTIYDRPV FCSLQSTSHQNPFLALVDHTTTFESGDAYGNLYVSGNHAFLEKDLQDLQHLAVG INSYNFMQLKAGATFQTPVELMVYTNKGLNAMSQAYHHLIRERVVRSEFKNGRPTV VNWTEATPDFNEALKLPVDEAKQLGIEHMFVLDGDFGHRDDNSLGDQVDRHKEF PQGLNHFVYIEHQGLGFWIGLEPEMISYDLSLYQQHPDYLMQVPGRSPSRNQXYIL DQGHVARNNIEDQLQKLSKOIDYIKWDMNRHLSDIYSVALPPERQGEVYHYVIG LYELLERTATYPHILFECGSGGGRFDAGMAYPMQIWASONTDAVARTIOGYTSL APYISLATAHVSQNOQTRGTSMSRASAASGVQIYELDTQLGSRADKOIVQKQV VOYKOIRPLQGESEYRKLSPITSNQAAKMFVSPQODEALVWVFNLTSAQPSLTKTK LVGLNPKLNYQNIATKAIIPGBDELQMGIFDPVVIYDITTKVTHFRAVTEN" 8265..8289 /standard_name="terminator 3" /note="putative" 9403..9426 /note="DR2; putative" /rpt_type=direct 10396..10442 /note="DR1; putative" /rpt_type=direct 10397..10443 /note="IR2; putative" /rpt_type=inverted 10681..10716 /standard_name="terminator 2" /note="putative" complement(10727..11593) /gene="scrK" complement(10727..11593) /gene="scrK" /function="fructose 6-phosphotransferase"	gene																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
------	--	------	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

Db	15324	ATGGCTAGTGGCGGTTACTTCTTGGCAATCCTCTTAGGATTTTCACGACGAAGCGC	15265	ORGANISM	Pediococcus pentosaceus
					Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae; Pediococcus.
QY	64	ttcggcggaatgagttcctgtggcgccgctatgttgatgagtggtgttcccagactt	123	REFERENCE	1 (bases 1 to 21839)
				AUTHORS	Leenhouts,K.K.J., Bolhuis,A., Kok,J.J. and Venema,G.G.
Db	15264	TTGCGTGGCAACCCCTATCTGGGTGCTAGG-ATGGGCATGATCATGGTCTTACCATCACT	15206	TITLE	The sucrose and raffinose operons of <i>Pediococcus pentosaceus</i> PPE1.0
				JOURNAL	Unpublished
QY	124	ggtgaacgctacgactgagcgccaccatgactcgcgccgaatgccaatggtgcctt	183	REFERENCE	2 (bases 1 to 21839)
				AUTHORS	Leenhouts,K.K.
Db	15205	AGTTAACGGCTATAGCGTTGCAACGACCATGGCAGCGCGCAAGATGGTTTACTTGGAAACGT	15146	TITLE	Direct Submission
				JOURNAL	Submitted (27-APR-1994) Leenhouts K. K., Institute for Biological Sciences, Department of Genetics, Kerklaan 30, Haren, The Netherlands, 9751 NN
QY	184	gtttggttttagatgtttgcccaagccggtttaccaggcgcacgtgtcttccctgtgtggtggt	243	COMMENT	On May 26, 1994 this sequence version replaced gi:475962.
				FEATURES	Location/Qualifiers
Db	15145	CTTTGGGTTACAGTTGTCACAAAGCGGCTATCAAGGCCAAGTGTCTCCAGTCTTAGCGGT	15086	source	1. .21839
					/organism="Pediococcus pentosaceus"
QY	244	ttcttggtatctggcaacgatcgagaaagtctctgcacagcgactcaagggccaactgcaga	303	repeat_unit	/strain="PPE1.0"
					/db_xref="taxon:1255"
Db	15085	CGCCTTCATCTAGCTACGCTTGAATAATCTTCCATAAACACATTAAGGGGCATTTGA	15026		487..510
					/note="DR2"
QY	304	cttctctgatacctccagtgactgaactgctgctcacccggtatccctacattcatcgccat	363		/citation=[1]
					/rpt_type=DIRECT
Db	15025	CTTCACGTTTACCCCGATGTTGGCCATTGTGATTACTGGTTTCTTAACATTTTACATCGT	14966	repeat_unit	488..561
					/note="IR1"
QY	364	tggccagcaatgcgtggtggggcgatgctggtgcacacggtctcacaggactttatga	423		/citation=[1]
					/rpt_type=INVERTED
Db	14965	TGTCGCCGCTTACGAACGCGTAGCGATGCAATTAACTAACGGTTTAGTGGCTTATATAA	14906	terminator	513..561
					/rpt_type=DIRECT
QY	424	tttcggtgtccagtcggcggtgtctctctgtgtgtgtctctactcaacatcgctcatcac	483	gene	2724..2760
					/standard_name="terminator 4"
Db	14905	CAGCACCGCTGGATTGGTATGGGAATCTTGGTTTATTATATCTTGTCAATTTGTTATTAC	14846	complement(2929..3864)	/citation=[1]
					/gene="rafr"
QY	484	tggtctgcaccagtccttcccgcaattgactggtgagctgtt-----taacca	531	complement(2929..3762)	/gene="rafr"
					/standard_name="regulator"
Db	14845	TGGTCTCCATCAAACTTTCCAGCAATCGAAACCCAGTTGTTGGCCAAATGTTGCCAAAC	14786	/function="regulation raffinose-operon"	/citation=[1]
					/codon_start=1
QY	532	gggtggtatcttcatcttcgcgaacggcatctatggtgtaataatcgccaggggtcgccatg	591	/product="regulator"	/transl_table=11
					/protein_id="CAA83663.1"
Db	14785	TGTTGGTTCTTTATCTCCAGTCGCCCTCGATGGCCAAATTTGGGCAAGGTGCGCCGAC	14726	/db_xref="GI:475963"	/db_xref="SWISS-PROT:P43465"
					/translation="MNGEYKTLANKSFESNVLFQGEACLPNTYKGNVRDSYVIHY
QY	592	tttggcagtggttcttctcgtggaagatgaaagctcaaggcctgcagggtgtcttcagg	651		IQEGKGTFAAANHPTATVLKAGDIFILPKGTCTCYQADNDQPKWYFVIGFSAGIRTEAM
					LSGSLAOKCYLRQVONGHIYADLSLYKVLHPIKNSLINDVLLGLIYRFLFYDLRLWY
Db	14725	TTTAGCTATTTCTTTGGCCACTAAGAGCCAAAGAAAGCAAGAACGCCCTCACTTCTTCAGCCGG	14666		PADATNKKVSTEQFNLAVSYLQENYSTGCTIMDLCHVNLNLSRSLYTLFKTHANTSP
					QKLLTKLLEDAKQRLSTSNNSVQSIANWVGKDSFTFSKAFKRYSGASPSYRKISG
QY	652	tgtctcgcgtgttcttggttattacgagcctgcgactcttcggtgtgaacctctgcctcg	711	-10_signal	complement(3837..3842)
					/gene="rafr"
Db	14665	GGTCTCAGCGTTACTCGGAATTACGGAACCTGCTATTTTCGGGGTCAACCTCAAGATGAA	14606	-35_signal	/citation=[1]
					complement(3859..3864)
QY	712	ctggcggcttcttcagtggtatcggttaccgcagctatcggtggcgctttgattgcactctt	771	gene	/gene="rafr"
					3909..5918
Db	14605	ATTTCATTTGCTTTTGCAGCGATTGCCCTCAGGAATTTGCTTCAAGTTTCTTAGGACTATT	14546	-35_signal	/gene="rafr"
					3909..3914
QY	772	taatatcaaggcaggttcgttggcgctgcagggttcttctgggtgtgtttcttctattgatgc	831	-10_signal	/gene="rafr"
					3932..3937
Db	14545	CCACGTTCTGTCGGTGGCGATGGTGGCATCAGTCATCGGATTCATTTCGATTCGCTTC	14486	CDS	3993..5918
					/gene="rafr"
QY	832	tccagataggttcattcttgggtgtgtcagttgttacctcttcttcacgtcgttcggcg	891		/standard_name="permease"
					/function="raffinose transport protein"
Db	14485	AAAGTCGATTCAGCATTCATGCTCAGTCTGTCATCTCGTTCGTCGTCGCATTTATTCC	14426		/citation=[1]
					/codon_start=1
QY	892	agcgattgattatggc	907		
Db	14425	AACCTTTATCTATGCC	14410		
RESULT	6				
PPSURFOP/c					
LOCUS	PPSURFOP	21839 bp	DNA	BCT	24-MAY-1994
DEFINITION	P.pentosaceus (PPE1.0)	sucrose and raffinose operons.			
ACCESSION	Z32771				
VERSION	Z32771.1	GI:493728			
KEYWORDS	alpha-galactosidase; alpha-glucosidase; enzyme Iabc; fructokinase; insertion element; permease; regulator; sucrose.				
SOURCE	Pediococcus pentosaceus.				



Db	15025	CTTCAGGTTTACCCGATGTTTGCCATTGCTGATCTACTGTTTCCTTAACATTTTACAACTCGT	14966
QY	364	tggcccaagcaatgcgtcgggtggcgatgctgctgcacacggtctctacagggactttatga	423
Db	14965	TGGTCCGCTCTTACGAACGGTGAGGATGCATTAACATAACGCTTGTAGTAGGCTTATATAA	14906
QY	424	tttcgggtgctccagtcggcggtcgtctcttcggtcgtgctgctactcacaatcgtctcac	483
Db	14905	CAGCACCGGCTGGATTGGTATGGGAATCTTTGGTTTATTATATCTGCAATTTGTTATTATC	14846
QY	484	tggtctgcaccagtcctctcccgccaattgagctggagctggtt-----taacca	531
Db	14845	TGGTCTCCATCAAACTTTCCCGACGAATCGAACCCAGTTGTTGGCCAAATGTTGCCAAAC	14786
QY	532	gggtggagctcttcattcttgcaacgcgcattctatgcttaataatcgccaggggtgcggcatg	591
Db	14785	TGGTGGTTCCTTTATCTTCCCAAGTCGCCCTCGATGCCAAACATTGGGCAAGGTGCCGCCAC	14726
QY	592	tttggcagtggtctccctggcgaagagtgaaaagctcaaggccctgcagggtgcttcagg	651
Db	14725	TTTAGTATTTTCTTTGGCCACTAAGAGCCAAAGCAAAAGCCCTGACATTTCTTACGCCCG	14666
QY	652	tgctcccgctgtcttggtattacggagcgtcgcatcttcggtgtgaaccttcgctgcg	711
Db	14665	GGTCTCAGCGTTACTCGGAATTACGGAACCTGCTATTTTCGGGGTCAACCTCAAGATGAA	14606
QY	712	ctggccggttcttcacgtgatcggtacgcagctatcggtggcgctttgatgcactctt	771
Db	14605	ATTTCATATTTGCTTTGACAGCATTCGCCATTCGCAAGAAATGCTTCAGCTTCTTTAGGACTAT	14546
QY	772	taataccaaggcagttgcgttggcgctgcaggttctctgggtgtgttctctattgatgc	831
Db	14545	CCAGTCTGTCGGTGGCATGGTCCCGATCATGTCATCGGATTCATTTCGATTCGATTC	14486
QY	832	tccagatagtgatgttcttgggtgtgctcagttgttaccttcttcacgtcgttcggcg	891
Db	14485	AAAGTCGATTCAGCATTCATGCTCAGTCTGTCATCTCGTTCGTGCGCATTTATTTC	14426
QY	892	agcgattgcttatgc	907
Db	14425	AACCTTTATCTATGCC	14410
RESULT 7			
LOCUS	LL297015	5800 bp	DNA
DEFINITION	Lactococcus lactis cremoris sucrose gene cluster.		
ACCESSION	297015	BCT	05-APR-1999
VERSION	297015.1	GI:4581474	
KEYWORDS	enzyme II sucrose protein; fructokinase; sacA gene; sacB gene; sacK gene; sacR gene; sucrose-6-phosphate hydrolase.		
SOURCE	Lactococcus lactis.		
ORGANISM	Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae; Lactococcus.		
REFERENCE	1 (bases 1 to 5800)		
AUTHORS	Luesink,E.J., Marugg,J.D., Kuipers,O.P. and de Vos,W.M.		
TITLE	Characterization of the divergent sacBK and sacAR operons, involved in sucrose utilization by Lactococcus lactis		
JOURNAL	J. Bacteriol. 181 (6), 1924-1926 (1999)		
MEDLINE	99173919		
REFERENCE	2 (bases 1 to 5800)		
AUTHORS	Luesink,E.J.		
TITLE	Direct Submission		
JOURNAL	Submitted (24-JUN-1997) Luesink E.J., Department of Biophysical Chemistry, Netherlands Institute for Dairy Research (NIZO), P.O. Box 20, Ede 6714 BA, THE NETHERLANDS		
FEATURES	Location/Qualifiers		
source	1..5800		
	/organism="Lactococcus lactis"		
	/sub_species="cremoris"		
	/db_xref="taxon:1358"		

gene	complement(330..1286)
CDS	/gene="sacA"
	complement(330..1286)
	/gene="sacB"
	/function="negative regulator"
	/codon_start=1
	/transl_table=11
	/product="SacB"
	/protein_id="CAB09692.1"
	/db_xref="GI:4581475"
	/translation="MIKLEDVANKAGVSVTTVSVRVNRKGYLSDATISKVKAMODLH YIPNAARSLOQSKLIGLVFPIKNIIFYAELEKIQALFIRGYKAMLATTEHDEQ KERDYLALLSNQVDGIYGHNLKAHDYIAEAPVAFDRLLTPTETVWSDNPEGG ILAKALINSKGKIIAFTGNDNTNSPTYLRRDGYLLELERNOLKPHIIKIPSOHTLL RKKEIKILLENDFDGVFCFDDDLTALLYKOLASNLKSLNVGDFGTETIENYPNL TTIQPINDLAEILLIRKIDGNDIITITQLPVLHIGID"
gene	complement(join(1283..2704,2721..2725,2737..2743,2761..2767))
CDS	/gene="sacA"
	complement(1283..2704)
	/gene="sacA"
	/EC_number="3.2.1.26"
	/codon_start=1
	/transl_table=11
	/product="sucrose-6-phosphate hydrolase"
	/protein_id="CAB09689.1"
	/db_xref="GI:4581476"
	/translation="MKWSTKORYTYDSYSESDLESRLKALKSPWKNFHLPEPTGL LNDNGFSYFNEKHLFYOHFFPGVHGLKSWHLVSDDLVHFETKGLVLPDTRIDN AGVYSGSALAFENFLIYTNHGRGEDWVRTPYOLGARIDKNQLVKFTEPLIYPDFS QTTDFRDPQIFSFQGIYCLIGASQSKNGIKLYKAIENNLTDKDLGDLDFSKEK MYMECPNLIPINGRSVLVFCPOGLDKSIVKDNIPYVVIADDFTTGSKNKLKNA GOLNLDGFCYATQSFNAPGSAIWSLGLPETSYPYTKYNQGVLSMVKKLSIK DNKLYQYFVKMKELRQEQDLLADNNIITSNYSLEYVDFRQQTSTILLSMTNEKGD SALKVEIDKENNTITLIRNYEKRLAHVKIERMNFIDQSIIFEIFINDGKVLSDCRVF PNKQYISIRSQNPRIKLWELKK"
RBS	complement(1294..1298)
gene	/gene="sacR"
	complement(1294..1298)
gene	/gene="sacR"
	complement(1390..1396)
	/gene="sacR"
	complement(1390..1396)
-10_signal	/gene="sacR"
	evidence=experimental
-35_signal	complement(1416..1422)
	/gene="sacR"
	evidence=experimental
gene	complement(1416..1422)
	/gene="sacR"
RBS	complement(2721..2725)
-10_signal	/gene="sacA"
	complement(2737..2743)
	/gene="sacA"
-35_signal	complement(2761..2767)
	/gene="sacA"
	evidence=experimental
gene	/gene="sacB"
	2815..4835
	evidence=experimental
	/gene="sacB"
-10_signal	2837..2843
	/gene="sacB"
RBS	2872..2877
	evidence=experimental
CDS	/gene="sacB"
	2883..4835
	/gene="sacB"
	/EC_number="2.7.1.69"
	/codon_start=1
	/transl_table=11

```
/product="enzyme II sucrose protein"
/protein_id="CAB09690.1"
/db_xref="GI:4581477"
/translation="MNHQVAERILNAVGRDNIQGRHCAIRLRLVLKDTGVTDQBAL
DNDPLDKGTFEAGVQIIVPGDVNTVVEEFIKLGIIEASTADLKEIAGSKQKNP
VMALVLLSDIFVPLIPALVAGLLMALNALTAEHLFATKSLVENFPMWKGFDIVN
TSAAPFTFMPILIGYSATKRFNGPYLGVAVGIMVMPGLINGYNVAEAINSHNTY
WDLFGKAOAGVOGVLPVIGVAFILAKRFLFKYLNDAIDFTFTPLLSVITGFL
TFTIVGPARLRFVNSGLTDGLVGLYNTLGAIGMLVFGFYSAIVVGLHQSFPALFPL
ITNYQHSGGDFIFPVAACANQAQATFALFVTKNKTALAPAGVSAIILGITE
PALFGNLKLPFFLALGASAIKSLFGLFVLAIVSLGSLGILGISIKRAGYNOQFM
ISIFSLFVSDVDFVSKIMGKGIKPKSGELFSPADGEIIIAVETGHAIGIKTNG
GEVLLHIGDITVSMNGNGFIONVKGVKAGDGLGSLGFKBEIKKSGLDOTVIIVTN
SASYNELPLSENVNDIKVGEKILLN"
RBS 4838..4842
/gene="sack"
4838..5720
/gene="sack"
4848..5720
/gene="sack"
/BC_number="2.7.1.4"
/codon_start=1
/transl_table=11
/product="fructokinase"
/protein_id="CAB09691.1"
/db_xref="GI:4581478"
/translation="MSVYVGSIEAGTKFVLAIADHFHNIKKFKFATTPQETISK
IKYKRNVSALGLSGFDIDNLNLSKTYITSTPKVGNKNLNGVKEALDIDPIY
FTVDNWSAYGEMKNTPKINLVLTIGTIGGGAIONGYFIIIGHSENGHQRINRH
DVLTFEGICPFHDCLEGAAGPSLEARTGILKEISDDPDWDLISYIAQAANAT
LTLAPCEIILGGVMEKPNKMSLIQKQFISMLNNYIDLPCSVKEYIRLPTVKENGSA
LGNFYLAISLFTKE"
BASE COUNT 1858 a 915 c 1031 g 1996 t
ORIGIN

Query Match 16.4%; Score 250.8; DB 1; Length 5800;
Best Local Similarity 51.1%; Pred. No. 7.2e-57;
Matches 725; Conservative 0; Mismatches 647; Indels 46; Gaps 4;

QY 4 atggcactcgcgcgtcgctgttgcagtggttgcagtggttgcagtcacccgcaacacgcgt 63
DB 3366 ATGAGTCCGCCACCATTTACCTTTATGCCAATTTTAAATAGGTATTCTGCCCAAAAAGA 3425

QY 64 ttcggcgcaatgagttcctggcgcgcgctattgtagcgatggttcccgagctt 123
DB 3426 TTGCGCGGAATCCATATTAGTGCGTG-TTGAGGTATGATTATGTTGATGCTCGACT 3484

QY 124 ggtgaacggctacgactggtgcccaccacatggtgcggcggaatgccaatgtggtccct 183
DB 3485 TATTAATGGATAAATGTTGCTGAAGCAATCTCTAATCATACAACTATTATGGGATAT 3544

QY 184 gtttggttagatgtgtcccaagcgggtaccagggcaccgctgcttctcgtgctggtgct 243
DB 3545 TTTTGGTTTAAAGTTGCCAGCGCTGGATATCAAGGACAAAGTCTCTCTCTAATGGAGT 3604

QY 244 tcttgattctgcaacgatcagaaagtctcctgcacagcgactcaagggcactgcaga 303
DB 3605 GGCCTTTATCCTTGCTTAACCTTGAGAGATCTTTTCATAAATACCTTAACGATGCTATAGA 3664

QY 304 ctctcctgatacctccagtgctgctgtgctgcacgggattccttacatcagcccat 363
DB 3665 CTTTCACATTTTACCTCGTACTTTTCAGTCATCATAACTGGATTTCTCACATTCATATTGT 3724

QY 364 tggccacgaacatgcgtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 423
DB 3725 GGGACACGACACTTCGTTTGTATCAAAATGGCTTGACAGATGTTGTTGGTGGACTTTATAA 3784

QY 424 ttcggtggtccagtcggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 483
DB 3785 CACATTAGGGGCACTAGGAATGCTTGTGTTTGGAGGCTTCTACTCGGCAATAGTAGTAAC 3844
```

```
QY 484 tggctgcaccagtccttcccgcaaatgagctgagctggttaaacag----- 532
DB 3845 TGAATTACATCAGAGTTTCCGGCCATTGAAACAATGTTAATCACAACTATCAACACAG 3904

QY 533 -----ggtggtatccttcctcgcacggcgtctatgctgaataatcccccaggtgc 585
DB 3905 TGGTATTGGTGGAGACTTTATTTTCCAGTTCGGCGCTCGCAAAATATGGCTCAAGCTGG 3964

QY 586 ggcattgttgagcagtggttctcctcgcgaagagtgaaagcctcaagggccttcgaggtgc 645
DB 3965 TGCACCTTTTGCATTTTATTTGTTACTAAAAACATTAAGACAAAAGCTCTGCAGCTCC 4024

QY 646 ttcagggtctcgcgtggttctggtattacggagcctgcgactctcgtggtggaaccttcg 705
DB 4025 AGCTGGTGATCTGCATTTCTAGGTATTACAGAACCTGCCTTATTTGGGATTATCTAAA 4084

QY 706 cctgcgtggcgcttcttcctcgcgtatcggtacccgagctatcggtggcgcttggattgc 765
DB 4085 ACTAAAAATATCCGTTCTTTATTTGCTCTGGGGCTTCAGCAATTTGTTCAATTATTATGG 4144

QY 766 actctttaatatcaaggcagtggttggcgcgctgaggttcttctggtgtgtttcttat 825
DB 4145 ATTATCCATGTCCTTGGGCTTAGTCTGGGATCCGAGGATTAATTTGGCTTTATCTCTAT 4204

QY 826 tgatgtccagatagtgatggttcttctggtggtgagttggttacccttcttcacgtcatt 885
DB 4205 TAAAGCTGGGTATAACTTACAATTTATGATTTTCGATATTTATTAGCTTTCTTATTCCTT 4264

QY 886 cggcgagcgatggttatggtccttacttgcgttccgcgcacgcgcagcattgatccaga 945
DB 4265 TGTGTTTACCTCAATCTATGCTGGCGGATGGAAGCTAATATCTATTACGAAAGAAAAA 4324

QY 946 tgaacccgctgctccagtgctgcaggaacgcaacgaagcgaagcgaagcaccgcaga 1005
DB 4325 TAAACAGATGC-----AACACTCAATACCAACCTCGAGAAATTTATTAT 4369

QY 1006 atttcaaacgattccaccatcatccaggcacctttgacccggtgaagctattgcactgag 1065
DB 4370 CGATCCAGTTAAAGTGGCGAACTCCTTGCTCCGATAAATGGATTTGTGATTCTCTGTC 4429

QY 1066 cagcgtcagcgatgcattgttccgcgcggaagcgttgcgtcggtggtggtccatcgtccc 1125
DB 4430 TGATGTAAGTGAACCTGTTTCTCAAAGAAATTTATGGGAAAAAGGTATTGCAATCAAGCC 4489

QY 1126 aaccaaggggcagtgatttctcgcgtgagtgaaagattggtggtcattcccatctgg 1185
DB 4490 TAAATCTGGAGAACTTTCTCACACGCCGCGAGAAATTTATTATTGCTTATGAACAGG 4549

QY 1186 ccatgcttccgagttcgcaccaaggtgaggtggttccaatgtgatatcttggatgca 1245
DB 4550 TCACGCTTATGGTATAAAAACAAAAATGGAGCT-----GAAGTTCTTTTACA 4597

QY 1246 cattggttcgacacagtaaacctcaacggcagcactttaaccgcctgaagcaggg 1305
DB 4598 TATTGGAATAGACTGTTTCAATGAATGTAATGATGATCATACAAATGTTTAAAGTTGG 4657

QY 1306 cgatgaagtcaacgagggagctgctggtgtagatcgatattggtggtcattgaagctgc 1365
DB 4658 CCAGAAAGTAAACGACAGGGATTTACTAGGATCTTTTGATAAAGAAATCAAGAAAGAG 4717

QY 1366 aggttatgaggttaaccacgcgcgattgtgttccgaatt 1403
DB 4718 TGGATTGGATGATGATGTAATTTATTGTTATTACAAATT 4755

RESULT 8
AE004395/10264 bp DNA BCT 31-JUL-2000
LOCUS vibrio cholerae chromosome II, section 52 of 93 of the complete
DEFINITION chromosome
ACCESSION AE004395 AE003853
VERSION AE004395.1 GI:9658068
KEYWORDS
```

SOURCE	Vibrio cholerae.	
ORGANISM	Vibrio cholerae	
REFERENCE	Bacteria: Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.	
AUTHORS	1 (bases 1 to 10264) Heidelberg, J.F., Eisen, J.A., Nelson, W.C., Clayton, R.A., Gwinn, M.L., Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Umayam, L., Gill, S.R., Nelson, K.E., Read, T.D., Tettelin, H., Richardson, D., Ermolaeva, M.D., Vamathevan, J., Bass, S., Qin, H., Dragoi, I., Sellers, P., McDonald, L., Utterback, T., Fleischmann, R.D., Nierman, W.C. and White, O.	
TITLE	DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae	
JOURNAL	Nature 406 (6795), 477-483 (2000)	
MEDLINE	20406833	
REFERENCE	2 (bases 1 to 10264)	
AUTHORS	Heidelberg, J.F., Eisen, J.A., Nelson, W.C., Clayton, R.A., Gwinn, M.L., Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Umayam, L., Gill, S.R., Nelson, K.E., Read, T.D., Tettelin, H., Richardson, D., Ermolaeva, M.D., Vamathevan, J., Bass, S., Qin, H., Dragoi, I., Sellers, P., McDonald, L., Utterback, T., Fleischmann, R.D., Nierman, W.C., White, O., Salzberg, S.L., Smith, H.O., Colwell, R.R., Mekalanos, J.J., Venter, J.C. and Fraser, C.M.	
TITLE	Direct Submission	
JOURNAL	Submitted (14-JUN-2000) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA	
FEATURES	Location/Qualifiers	
source	1..10264 /organism="Vibrio cholerae" /strain="N16961" /serotype="O1" /db_xref="taxon:666" /chromosome="II" /note="biotype: El Tor" complement(114..1553) /gene="VCA0653" complement(114..1553) /gene="VCA0653" /note="similar to GB:M76768 SP:P22825 PID:155262; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="PTS system, sucrose-specific IIBC component" /protein_id="AAF96554.1" /db_xref="GI:9658069" /translation="WDYPIAKLLESIGGKSNIOALAHCASTRRLRLVNDFTQINESA IESLQGVQGFQVACQYQIIFGSGIVNQYIAEMAKLTGLVEMSTNDVASAGAEKQNW QAVKGLSDIFVRIIPAIYVAGGLMLFNVLTPAGLFIQGLSIDANPGLADLASMIN TFANAPFYLPVLLAFSASRKFPGNPLGALMLVHPDLNLGNGFGSGASVSGTPT WNILGFEIEKYGQSVLPVLVSAYILAKIENGLRKIVPSVVDNLLTPMLAIFITGFL TPTVVGPLTRDVGMLGALNWLXPSAGFVGGLRGFTYAPVITGMHHSFTAIEGTL LADIVTGTGTFIEPIAAMSNIAQGAALAVGVTETKLKGVAIPSGVTALLGITEPA HGVNLKURYPFPIAICGAALASAFITLFNKQAQALGAAGLPGIISINPQQIGYYIMG MAISVPAFALTVLLAMREKTKQAQVTA" 1854..2846 /gene="VCA0654" 1854..2846 /gene="VCA0654" /note="similar to GB:L14685 PID:289864; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="sucrose operon repressor ScrR, putative" /protein_id="AAF96555.1" /db_xref="GI:9658070"	<p>/gene="VCA0653" /note="similar to GB:M74035 PID:155269; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="sucrose-6-phosphate dehydrogenase" /protein_id="AAF96556.1" /db_xref="GI:9658071" /translation="MLDTLLEAGGINNTRILAPQGVVLAKHPPLVPHLPDDVS LQSVLGEWQLSVORTAEVSDQQLAAIGKAIEROKLETLPYTALDCPYRPLWHISPP QGLLDPNGFIYHQGEYHLFYQWHPFACEHKDKYVWHLKSLDLVDWQWQSVALTPSDW LPEGVFSGHVSQHDQDLWLFYTGNTRLGVDRQRTMQCAARNANGFEKILGPVIRC DYGVMECPDPELOGEAFFVPGOGLASANPHHTIEHONRIFRATONAGOGIALQOG WPLDSGFDEYAPQATAGDRVLGWMGLPDETQHPSCDQGWILHQLTALRELVREG RIYQHPRELDTLQSEPHLLLSNDVTELTKSFALQVLTLPWGCSRLRLMONTQYRVTL TLDANQLRLDRSATQIRQDGTIRELKDSTPVELRLILADQSSLEIFINQGEHVMTS RIFTPTDASGSLHGSADAKLYMAPASAPFNLEVNVP" complement(4690..5661) /gene="VCA0656" complement(4690..5661) /gene="VCA0656" /note="similar to GB:M76768 SP:P22824 PID:155263; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="fructokinase" /protein_id="AAF96557.1" /db_xref="GI:9658072" /translation="MNALVRLSSNHLRSDSMSRVMLTGDVVDLTPDQOQHYLKCPG GAPANVAVALRCLGRSAFFRGVNDPFRFEMQOQLTDEQVDCQHLYPDPVHRTSTVY VDLDEHGERSTFMVKPSADQFLQISDIPSFONGEMLHVCSTALANQPSRSTFAAIA QMKVGGYVSFDPNLRREVSWPEPQELQATVMRAVLADVVKESSEELQFLTQTSIEE GLQATADFOIPLVVVTLGKAGALVVPNSQIVSGKAVKPIDTTTGADGAFVGLLLRL SVAQDHNHQATILDVAVKWANGCCGALATTQKGAMTALPNOAALYAFLE" 6008..7546 /gene="VCA0657" 6008..7546 /gene="VCA0657" /note="similar to SP:P13035 GB:M21277 GB:M55989 GB:M96795 PID:146181; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="aerobic glycerol-3-phosphate dehydrogenase" /protein_id="AAF96558.1" /db_xref="GI:9658073" /translation="MNAKSTRFTQDEVLDLIIVGGINGAGIAADATGRLKVLGYDA KDFASATSSAKFLVHGGLRYLEHYEFRLVSEALAEVLLKKAPHIVTPMRFLPHR PFLRPMWIRAGLFLYDLGKRTSLPASHKVNKAGSVTKPEMOIGFEYSDCWDDAR LVILNMQAEOGAELVNLCTVEKAERMGDLWHVTLDDQTOQRFERRSHALVNAAGP WVKQFLNNAHVSSPYGRLIOGSHIIVPRIHDEPQAVILQNEDKRIVEIPLYDDYS MIGTIDVETKGDPRKVAITDAERDYLISIVNKHFMREIARSDIIAEFSVRLCDDES NSPQAITRDYTLSDQDAEAPLLSIFGKLTTRYKLGAEAMKHLAPFPFKMKAQWTA DAPLPGENFDYALKNLQVAAFPFITESVIERWLSRYSRRTTOLLAGVTGIEDGIA FSGELYQREIDYLCEKEFARHAQDIFWRRSKLGNHDTSVVEVESYLQOKFHAEQPL KATV" complement(7648..9258) /gene="VCA0658" complement(7648..9258) /gene="VCA0658" /note="similar to GB:X61367 PID:48195; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="methyl-accepting chemotaxis protein" /protein_id="AAF96559.1" /db_xref="GI:9658074" /translation="MYTFALIKITLEQQAEMSAYTPSAQOEVLVGDHDLVSTDLKG VITYCNDIFCRITAGFOADELLGNINVRHASKRPAKPAFADMMHHLKQGHANRGIYKNR TKSGGFYWDAYVTYQQGQLTGYOSVRVKAERKWEIATIKAYOLAAEKAGAKIKQ FKLHTSLRYALLGALMPALGAQPEQWQWLASLLPAGVLGFLFROELVRYTPQOL KOWQNEYDSISRLIYSGADAFSVADYHLKMASARTILGRMWDSPARPLGELANQLHL TTQEVHQALAAQNSNIQAVTQATDAVESAAERVSSHTSHSAHLIDQVQDCAETKHSI</p>

```
NVTHQLQRLATQAESAAITLLKLSDAQVQOLMTEIGGIAEQTLNLLALNAIEAAR
AGSGRGFAVADVEALSGARTQOIQTSIDTLMSTIEAWRGDITARSQDTECA
QDANTTLOLOQDVECVMSDLRLVIGEVASAOHQRELTCEVNOHIHSIASVATNSAA
THVVEQLAMMSKVAEFGALSQFAQK"
complement(9444..10067)
/genes="VCA0659"
CDS
complement(9444..10067)
/genes="VCA0659"
/notes="similar to PID:2062640; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="protein F-related protein"
/protein_id="AAF96560.1"
/db_xref="GI:9658075"
/translation="MPEMFVWLCVTFLLTGGSLVDRMFNDNLLDVAPKGFDFVRY
PEWYAPQNTSVMPNMTOPRSASYEDLRQYLLSNGIQHEVYVPGDYPMKLNNTVRF
ETGSAVMSKQWLDIVARFATESGIDIVLEGHTDNTGSEKLNKLAERRANAVKA
ALVQSRVAQNAIYTRFGENVFACNTNSTRNGRACNRVREIRFLASN"
complement(10068..10190)
/genes="VCA0660"
CDS
complement(10068..10190)
/genes="VCA0660"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF96561.1"
/db_xref="GI:9658076"
/translation="MNGEHSNHLGDKGPIKLLMLTDKVFERNMPINYYKFNPN"
BASE COUNT 2711 a 2569 c 2371 g 2613 t
ORIGIN

Query Match 15.2%; Score 231.8; DB 1; Length 10264;
Best Local Similarity 55.4%; Pred. No. 1e-51;
Matches 497; Conservative 0; Mismatches 387; Indels 13; Gaps 2;

Qy 3 catggcatctgcgcgtcttcgttcgagctgttggtggtttccaccgcaaccaagcg 62
Db 1071 CTTGCGCAATGCCCTTTGTTTATTGGCTGTATTGTTGGCATTTTCTGCCCTCGCGCAA 1012

Qy 63 tttagcggaatagttctctggcgccgctgattggtatggcagtggttcccgagct 122
Db 1011 GTTTGGCGGTAAACCCCTTTTGGCGCGTGGC-CTGGGTATGCTGATGGTGCATCCCGATC 953

Qy 123 tggtagcgtacgacgtgcgcgcacacatgctgcggcgcaaatgcgaatggtccc 182
Db 952 TCTTGAACGGCTGGGGTTTGGGAGTGTCATCCCGTATCTGGCACCGTTCCAACTTGAATA 893

Qy 183 tgttggttttagatgttgcccaagccggttaccaggcacctgcttcctctgctggtg 242
Db 892 TTCTTGTTTGAATCGAAAGTGGTATTCAAGGCTCGGTATTGCCGCTGCTGCTGTTT 833

Qy 243 ttcttgattctggaacgactgagaagtctctgcacaaagcactcaaggcactgcag 302
Db 832 CCGCTTACATCTGGCCAAAATTTGAAATGGCTTACGCAAAATCGTACCGTCTGTAGTGG 773

Qy 303 acttctgatcaactcaactgactgactgctgctcaccggattccctacattcatcgcca 362
Db 772 ATAACCTGCTTACCCCAATGCTGGCATTTTATATCAGGGGTTTTCGACCTTTACTGTGG 713

Qy 363 ttggccacgaatgcctggtggcgatgctgcgcacacgctctacaggagactttatg 422
Db 712 TTGACCGCTAACCGGTATGTCGGTTCATGTTGGAGATGCACTGAACCTGGCTGTATG 653

Qy 423 atttcggtggtccagtcgcgcggtctctctcttcggtctggttactaccacaaatcgatca 482
Db 652 ACTCGCAGGGTTGTCGGTGGCGCACTGTTTGGCTTTATCATGCTCTCTTTTGTGATCA 593

Qy 483 ctggtctgacacagtccttcccgcaaatgagctgagctggtt-----aacc 530
Db 592 CCGGAATGCACCATAGCTTTATCCGATAGAAACTCAGCTACTTGCCTATGCGGATATCGTGACCA 533
```

```
Qy 531 aggtggtatcttcattctgcgaacgcatctatggtcctaataatcgccagggtcgcgcat 590
Db 532 CGGTTGCACCTTTATTTCCCTATCGCTCGATGTCTAACTTGCAGAGGCGCCGCGG 473

Qy 591 gtttggcagtggtctctctgcgaagagtgaaagctcaaggcgcttgcagggtgcttcag 650
Db 472 CTTTAGCTGTGGGTGTGATGACGAAAGAGACCAAACTCAAAGGGTGCAGATTCCTTCGG 413

Qy 651 gttctctcgtgttcttgattacggaagcctgcgactcttcgtgtggaaccttcgctgc 710
Db 412 GTGTGACCGCGCTGCTGGGATCACTGAACCTGCCATGTTTGGCGTGAATTTGAAATGC 353

Qy 711 gctggcgcttcttcacgtggtaccgcagctacgtgcgccttcatttcacgtcctct 770
Db 352 GTTATCCGTTTATGCTGCCATCTGTGGTGGCTTTTAGCCAGTGCCTTTCATCATTAT 293

Qy 771 ttaatacaaggcagttgctggcgctgcaggttcttctgggtgttggtttcttcatgatg 830
Db 292 TCAATGTAAAGCCCAAGCGCTAGGCGCGCTGTTTGCAGGATCATTTTCGATTAATC 233

Qy 831 ctccagatagtgatggttcttgggtggtgcagttgttaccttcttcacgtccttcg 887
Db 232 CACAGCAAAATGGTTTACTACATCATGGGTATGGCATCTCATTTGTGGCGCCTTTG 176
```

## RESULT 9

```
STRSCRA STRSCRA 2508 bp DNA BCT 21-MAR-2000
LOCUS Streptococcus mutans sucrose-6-phosphate hydrolase (scrB) gene,
partial cds; and enzyme scr-II gene, complete cds.
DEFINITION M22711
ACCESSION M22711 GI:153799
VERSION 1
KEYWORDS enzyme scr-II; phosphoenolpyruvate-dependent sucrose
phosphotransferase system; sucrose-6-phosphate hydrolase.
SOURCE Streptococcus mutans.
ORGANISM Streptococcus mutans.
Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
Streptococcus.
REFERENCE 1 (bases 1 to 2508)
AUTHORS Sato,Y., Poy,F., Jacobson,G.R. and Kuramitsu,H.K.
TITLE Characterization and sequence analysis of the scrA gene encoding
enzyme IIsr of the Streptococcus mutans phosphoenolpyruvate-
dependent sucrose phosphotransferase system
J. Bacteriol. 171, 263-271 (1989)
JOURNAL 89123027
MEDLINE
COMMENT Draft entry and computer-readable sequence for [1] kindly provided
by H.K.Kuramitsu, 22-FEB-1989.
FEATURES
source
1..2508
/organism="Streptococcus mutans"
/strain="GS-5"
/db_xref="taxon:1309"
/clone="pMH613"
/complement(<1..28)
/notes="sucrose-6-phosphate hydrolase (scrB)"
/codon_start=1
/transl_table=11
/protein_id="AAA26970.2"
/db_xref="GI:7272364"
/translation="MNLQNIRY"
223..2217
/notes="enzyme scr-II"
/codon_start=1
/transl_table=11
/protein_id="AAA26971.1"
/db_xref="GI:153801"
/translation="MDYSKVASVITAVGKDNLVAAAHCACTRLRLVKDSDKOKAL
DKNADYKGTFTDGOYQVLTIGPDYVYDEIITKGTLEVTSTDLKTAASGKFNK
IMALKLSDIFVPIIPALVAGGLMALNFLTSGLFSTKSLVQGFPIIKGSSDMIQ
IMLAAFPWFLPILVGISAKRFGANQFISGIMVAPGAANIIGLAANAPISKAAT
IGATGFWNIIFGLHVTQASYTVQVILVAVWLLSILEKFFKRLPSAVDFITFTPLLS
VIITGLFTFIVIGPVNKEYSDWLTNGIVNLYDTTGLGMGVEGALYSPVYMTGLHQSF
```



```
PAIETOLISAFONGTGHGDFIVTASMANVAQGAATFAIYFLTKDKKMKGLSSSGVS
ALLGITPALGVNLKYPFEPFCALIGSASAAAGLLOVVAVSLGSAGFGFLSIIKA
SSIPFVVCLELISFAIPAYYGGKTKAVDFRAEAAREAEIEVQPIPEASAAAN
KAQTVDEVLAAPLAGEAVELVSVDVPSSEMGKGAIKPSGNTVAVPDVGTVOIAF
DTHAYGKISONGAEIILHIGIVDSMEGKGFQKQVADQIKRKGDLVLTGTFDSKIAE
AGLONTMFIYVNTADYASVETLASSGTAVAGDSLSLEVKK"
misc_signal
2227..2252
BASE COUNT 740 a 401 c 513 g 854 t
ORIGIN
Query Match 15.1%; Score 230.8; DB 1; Length 2508;
Best Local Similarity 50.1%; Pred No. 1.9e-51;
Matches 779; Conservative 0; Mismatches 717; Indels 58; Gaps 6;
QY 2 teatggcatctgcgccttcgcttcttgcagtggttgcagtggttgcagcaaccaagc 61
Db 704 TAATGTCAGCAGCGCTTCTGCTTCTGCCCATTTTGGTGGGATTTTCAGCGGCTAAAC 763
QY 62 gtttcggcggaatagtttcctggcgccgcgctattggtatgtagcgatggttcccgagc 121
Db 764 GTTTGGTGCTAACCAATTTCTAGGAGCTCA-ATCGGCATGATTATGGTCGCACCGGG 822
QY 122 ttggtgaacggctacgacgtggcgcca-----ccatggcgtgcg 160
Db 823 GCAGCAATATCATTTGTTTACAGCTAATGCTCCAAATTTCAAGGCTGCTACTATTGT 882
QY 161 ggcgaatgccaatgtggtccctgttgggttagatgttgcgaagcggttaccaggcg 220
Db 883 GCTTATACAGGATTTTGAATATTTTGGTGGCTGCTTACTCAGGCAAGTTTATACTTAT 942
QY 221 accgtcttcctgctggtgttcttgattcttggaatctggcaacgatcgagaagttcctgcac 280
Db 943 CAAGTGATTCAGTACTTGTGCTGTATGGCTTTTATCTATCTTAGAAAAATTTTCCAT 1002
QY 281 aagcgactcaaggcgactgcagactcctgacacccagtgctgacgtgtgctgcacc 340
Db 1003 AAGAGATTCCTCAGCAGTTGACCTTACCTTTACACCTTTATGTCAGTTATTTACT 1062
QY 341 ggattcctacattcatccatggccagcaatgcgtggtgggagatgctggca 400
Db 1063 GGATTTTTCACATTTATCGTTTATGCTGCTGTTATGAAGAAGATTTCTGACTGGCTTACA 1122
QY 401 cagcgctcacaggactttatgatttcggtgtccagtcgcggtctgctcttcgctcg 460
Db 1123 ATGGNAATCGTATGCTGTATGATACACAGGATTCCTAGGTATGGGTGCTTTTGTGTCT 1182
QY 461 gtctactaccacatcgatcactgctgctgcacagtccttcccgccaattgagctggag 520
Db 1183 CTGTATTCACCTGTAGTTATGACTGCTCTCCATCAAAAGTTTCCAGCTATTGAAACCCAA 1242
QY 521 ctggt-----taaccagggtgataccttcattcttcacagcgcatct 562
Db 1243 CTCATTTACAGCTTTCCAAAATGGTACTGTGCTATGTGACTTTTCTTTTACAGCTTCA 1302
QY 563 atggcctaatacgccagggtgcgcagatgtttggcagtgattcttcctggcgaagatgaa 622
Db 1303 ATGGCAATGTCGCACAGCGCAGCAACCTTTGCANITTTACTTTTAAACAAAGATAAG 1362
QY 623 aagctcaaggcgcttcgagtgctgcagtgctcctcgctgttcttggtattaccggagcct 682
Db 1363 AAGATGAAGGCTCTCATCTTCTTCAGGTGTGTCAGCGCTTCTGGTATTACAGACCT 1422
QY 683 gcgattcctgggtgaacctcgcgtgcgttcttccttccttcacggtatccgcga 742
Db 1423 GCTCTTTTGGGGTGAATTTGAATATCCTTTCCCATTTCTTTTGTGCTCTGATGGCTCA 1482
QY 743 gctatcggtggcgcttgatgactcttataatacaaggcaagtgcgttggggcgctgca 802
Db 1483 GCTAGGCTGCAGCAATTCGGGGCTTACTTCAAGTGTAGCGGTTTCTCTTGGTTCGGCA 1542
QY 803 ggtttcttgggtgtgttcttctatgatgctccagatagtgatggttcttgggtgca 862
```

```
Db 1543 GGTTCCTTGGTTTCCCTTCTATTAAAGCAAGTCTTATTCATTTTATGCTCTGTGAA 1602
QY 863 gttgttaacttcttccttcacgcagcagcaggttcttccttccttccttccttccttc 922
Db 1603 TTAATCAGTTTGTCTATAGCTTTTGCAGTAACCTTATGGCTATGCTTAAGACGAAGCGTGT 1662
QY 923 cgcacaggcagcatgcatgcagatgcaaccgctgctccagtcgctgcaggaacgacaaa 982
Db 1663 GATGCTTTTGGCTGCAAGCTGCTGCGAAGAGCTATTGAAGAAGTGCAAGAAATTCGA 1722
QY 983 gccgaagcagaagcaccgcagcaaatcttcaaacgattcca---ccatcatccaggcacct 1039
Db 1723 GAAGAAGCTGCTTTCGACCAATAAAGGCACAAAGTTACTGATGAAGTCTTTCGGGCTCT 1782
QY 1040 ttgaccggtgaagcttcttcactgcagcgcgtcagcgcgtggttccttccttccttccttc 1099
Db 1783 CTGTGCTGTTGAAGCTGTTGAATTAACCTCTGTTAATAGCCCTGTTTCTTCTAGCGAAGA 1842
QY 1100 ctggtcgcggcggttgcctcgtcccaacaaaggcgaggttagtttcttccttccttccttc 1159
Db 1843 ATGGGTAAAGGGATTTGCTATCAAACTAGTGTATATACAGTTTATGCACCGGTTGATGG 1902
QY 1160 aagattgtggtgccttcccatctgccaatgcttccgagttccgacccaagcgtcaggat 1219
Db 1903 ACTGTTCAAAATGCTTTTGATATCTGCGCATGCTTATGGTATTAAAGTCACACAATG --- 1957
QY 1220 gttccaatgtgatatcttctgacacatggttcttcgacacagtaaacctcaacgcagc 1279
Db 1958 -----GTGCGGAGATTCTTATTATCATTTGTTATGATACAGTATCAATGGAAGGTAAA 2010
QY 1280 cactttaaccgctgaagaagcaggcgcgtgaagtcagaagcagcagcagcgtcgtgtgaa 1339
Db 2011 GGATTTGAACAAAAAGTTCAAGCAGATCAAAAAATTTAAAAAGGTGACGCTTCTCGGAACA 2070
QY 1340 ttcgatattgctgcttaaggctgaggtatgaggttaaccacgcgcgattgttcttcg 1399
Db 2071 TTTGATAGTGACAAATTCGACAGCTGCTCTTGATAATACAAATGTTTCATTGTTACT 2130
QY 1400 aattacaagaacacgcgcgtgtaaacattcaggttcttcggaattgaaagcggagcc 1459
Db 2131 AATACAGCTGATTATGCTGCTGAGTGAACCCCTCGCTTCTTTCAGGAACCTGTTGCTGTAGGT 2190
QY 1460 ---aacctgctcaacgctcgaagaagaagcgggtgctccagcaacacccaagct 1510
Db 2191 GATAGCTTACTTGAAGTTAAAAAATTAAGAAATATTATCAGAAAGACCGCTAAGGT 2244
```

```
RESULT 10
AE006222/c
LOCUS Pasteurella multocida PM70 section 189 of 204 of the complete genome.
DEFINITION Pasteurella multocida.
ACCESSION AE006222.1 GI:12722266
VERSION AE006222.1
KEYWORDS
SOURCE Pasteurella multocida.
ORGANISM Pasteurella multocida.
REFERENCE 1 (bases 1 to 10085)
AUTHORS May, B.J., Zhang, Q., Li, L.L., Paustian, M.L., Whittam, T.S. and Kapur, V.
TITLE Complete genomic sequence of Pasteurella multocida, PM70
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 98 (6), 3460-3465 (2001)
MEDLINE 21145866
REFERENCE 2 (bases 1 to 10085)
AUTHORS Zhang, Q. and Kapur, V.
TITLE Direct Submission
JOURNAL Submitted (24-OCT-2000) Department of Veterinary Pathobiology, University of Minnesota, 1971 Commonwealth Ave., St. Paul, MN 55108, USA
FEATURES
Location/Qualifiers
```





```
QY 305 ttctgtactcactcgaagtgtgagctgtctgctcaacgagattccctacattcattcgcatt 364
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1176 AACCTTTAAACGCCATTACTAGCGATTTTCATAGCGGTTTGTAGCGTTTACTGTGCTC 1235

QY 365 ggcagcaatcgctggtggcgatgctgctggcacacgctgtacaggacatttatgat 424
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1236 GGTCGGTTCACGGGTGATATCGGTTTCCATATGGGTGATGCTCTGAACCTGGCTGTACAAC 1295

QY 425 ttggtgtccagctgcggctgctgctgctgctgctgctgctgctgctgctgctgctcatt 484
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1296 ACAGCTGGCTTTGTGTGGTGGCTGATTTGGCTTATTTATGCTCCGTTGTATTTACT 1355

QY 485 ggtgtgcacagctcctccgcgaattgagctgagctggtt-----aacag 532
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1356 GGTATGATCACAGCTTCAATTGGCATTTGAACCTCAACTTCTTGGGATATTGCAACGACG 1415

QY 533 ggtggtatccttccttcgcaacggaatctatggttaataatcgcccaaggtgcggaatgt 592
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1416 GCGGGCACCTTTATCTTCTTCTATTTGCGGGCATGCTTAACGTTTCTCAAGGAGCGGCAGC 1475

QY 593 ttggcagctgtcttcctggtgcaagagtgaaagctcaaggccttcagctgcttcaggt 652
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1476 TTGGCGTGGCGGTGATGACAAAGATAAGAGATGAAGATATGGCATTCATTCATCGGT 1535

QY 653 gtctcogctgttcttggattacggagcctgcgatcttcggtggaaccttcgcctgcgc 712
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1536 GTTACGGGTTTGGTGGGTATTACAGAACACGACCATGTTGCTGTTAACTTAAACACTACGT 1595

QY 713 tgcagcttcttcagctatcggtaccgcaagctatcggtgagcttgccttcattt 772
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1596 TACCATTTTATTCGGCGAGTTTTCGGCTGGCGGCTTATCAAGTGGCTTTCATCACAAGTTT 1655

QY 773 aatcaagcagctgtcgttggcgctgcaggttcttcgtggtgttcttattgatgct 832
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1656 AATGTGAACCTCAGCATTTGGTGGCAGCTGGCTTCCAGCGATCATTTCCATCACGCCA 1715

QY 833 ccagatatgctatgttcttgggtgtgagttgttacccttccttcacgcatcgcgca 892
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1716 GATAAGATTGGGTACTACATCGGGGTATGTTGATCGCGTCTCTTGACCGCATTTGTTCTC 1775

QY 893 gcatgcttattg 906
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1776 ACCATTGTTCTGG 1789
```

```
RESULT 12
SXSCRA          2655 bp  DNA      BCT      10-FEB-1999
LOCUS           S.xylosus scrA gene and unidentified open reading frames.
DEFINITION      X69800
ACCESSION       X69800
VERSION         GI:407905
KEYWORDS        membrane protein; scrA gene; sucrose transport protein.
SOURCE          Staphylococcus xylosus.
ORGANISM        Staphylococcus xylosus.
REFERENCE       Bacillus/Firmicutes; Bacillus/Clostridium group;
AUTHORS         Bacillus/Staphylococcus group; Staphylococcus.
TITLE           1 (bases 1 to 2655)
JOURNAL         Direct Submission
                Submitted (15-DEC-1992) E. Wagner, Mikrobielle Genetik,
                Universitaet Tuebingen, Auf der Morgenstelle 28, 7400 Tuebingen 1,
                FRG
                2 (bases 1 to 2655)
                Wagner,E., Gotz,F. and Bruckner,R.
                Cloning and characterization of the scrA gene encoding the
                sucrose-specific Enzyme II of the phosphotransferase system from
                Staphylococcus xylosus
                Mol. Gen. Genet. 241 (1-2), 33-41 (1993)
FEATURES        94049686
                Location/Qualifiers
                1..2655
                /organism="Staphylococcus xylosus"
                /strain="DSM 20267"
```

```
/isolate="cured strain C2a"
/db_xref="taxon:1288"
/clone_lib="sub-genomic in Staphylococcus xylosus mutant
scr1"
/clone="pSE4"
complement(<1..121)
/note="ORF3"
/codon_start=1
/transl_table=11
/protein_id="CAA49459.1"
/db_xref="GI:407906"
/translation="SPTREMBL:Q56203"
complement(132..137)
338..342
350..925
/note="ORF2"
/codon_start=1
/transl_table=11
/protein_id="CAA49460.1"
/db_xref="GI:407907"
/translation="WOHIYSEFYNQESKTKLKSVLGIVLHDDVNNNSAEDYIVWLQ
QRIDNLEKKGWACAYVDKQTCYFHPSPYVEWHGNAFANDHYIGIERCQSKINGIL
TDQFGNEEASWIAALLKKYQLPVNRETVKLHKMFEDTECPARAWSIHLDNAPT
EENIKMLQDYFDIKIKYHVFNIKESDMFVG"
972..977
995..1000
1040..1045
1053..2495
/gene="scrA"
1053..2495
/gene="scrA"
/note="ORF2"
/codon_start=1
/transl_table=11
/product="EiIsr"
/protein_id="CAA49461.1"
/db_xref="GI:407908"
/db_xref="SWISS-PROT:P51184"
/translation="MNYKSAENILQALGEDNVEAMTHCATRLRLVLKDEGLVDEKA
LGMDDVKGTFSTGGQVITIGSTGVKVFSELEKITGREASSVEKTKGNTKNNPF
IRFVKMLSDIFVPITPAIVAGGLGMINNITAPFYDNOSLIEYQNOFSGLAEMIN
IFANAPFTLLPILIGFSAKRFGGNAYLGAALGMILVHPDMSAYDYPKALEAGKEIP
HWNLFGLINOVYQGVPLMVAIVLATIEKLRKVIPTVLDNLTPLAILSTGF
ITPFSVGPLTRTLGWSLGLTWLVEFGAIGLIFGLLYAPIVITGMHHSFAIETQ
LIADSSSTGSGFPIFPATMSNINQGAALAAAFIIRKNNKLGKVASAGVSALLGITE
PAMFGVNLRLRYPFICAIVSGSGISAYIAFFKVKALATAGTIPGFISISGQNNGLH
YGIAMIIAFIVAFVGYTALSIRKKYRNIEA"
BASE COUNT 909 a 347 c 521 g 878 t
ORIGIN
```

```
Query Match 13.8%; Score 210.6; DB 1; Length 2655;
Best Local Similarity 56.2%; Pred. No. 5.6e-46;
Matches 485; Conservative 0; Mismatches 359; Indels 19; Gaps 4;

QY 5 tggcatctgcgcgttcgcttcttcgcatgttgcagttggttgcagcgaaccagcgtt 64
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1537 TTGCTAATGCCCATTTACATTTATTCCTATACTAATTTGTTTGTGCTGCAAAACGAT 1596

QY 65 tcggcgcaatgagttcttcggcgccgcgttatggcgatggcggtgttcccgagcgtt 124
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1597 TTGAGGCAATGCATACCTCGAGCGGCAC-TTGAATGATATTGTTTCATCCAGAAATTG 1655

QY 125 gtaacggtctacgactggtgcgcacacatggtgcggcgcaa---atgcaaatgtgtgtc 181
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1656 ATGATGCTTATGATTCATCTAAAGCATTAAGACGCGGAAAGAGATTCCTCATTTGGAAC 1715

QY 182 ctgttgggttagatgttgcgaagcgggttaccaggaccgttcttcctgtgctgtg 241
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1716 TTATTCGGTCTCGAAATTAATCAAGTAGGTAGGTACCAAGGTCAAGTCTTACCAATGCTAGTA 1775
```

QY	242	gtttcttgatctgcgaacgatcagaagtcttcgtcacacagcgactcaaggcactgca	301
Db	1776	GCTACATATATCTTGCACAATGTAAGAAAGCTAAGAAAGTTATACACAGTATTA	1835
QY	302	gacttctgatcactccagtcgactgctgctgcctcaccggatctcctacattcatcgcc	361
Db	1836	GATAATTTATTAACACCTTTATTAGCAATATTATCAACTGGCTTTATAACATTTTCATT	1895
QY	362	attgcccagcaaatcgctgggtggtgcgctgctgctggtgcacacggctacagggactttat	421
Db	1896	GTTGGTCCATTAAACAGTACTTATAGGGTATTGGTTATATCAGACGGTTTAAACGTGGCTATAT	1955
QY	422	gatttcggtggtccagtcgagcgctgctctctcgctgctgctgctactaccacaaatcgctcatc	481
Db	1956	GAATTTGGTGGTCTATCGGTGGCTTATTTGGTTGTTATATCGGCCAATCGTTATT	2015
QY	482	atgtgtctgcaccagtcctcccgcaattgcagctggagctggttaacca-----	531
Db	2016	ACAGGTATGCATCATAGATTTTATTGCTATTGAAACGCAACTTATCGTCGACAGTTTCATCA	2075
QY	532	--gggtgatacttcattctcacaacggcatctatggttaataatcgccacagggtgcgca	589
Db	2076	ACGGGTGGTTCATTATATTTCCCTATTGCAACAATGCTTAATATTCACACAGGAGCGGCT	2135
QY	590	tgttgccagctgtctctctcctgcgca---agagtgaaggctcaaggccttcgagtgct	646
Db	2136	GGCTTGCAGCGTCTTTTATTATAAAGAAATAAATAATTAAGAGCGTAGCGCTCTGCT	2195
QY	647	tcaggtgtctcgcgtgtcttctggtattacggagcctgcgactctcggtgtaaaccttcgc	706
Db	2196	GCAGGTGTATCAGCTCTATTAGGTATTACAGAACCAACGAATGTTTGGTGTGAACCTTGAAG	2255
QY	707	ctgcgctggcgtctcttcacatcggtatcggtaccgacgactatcggtggcctttgattgca	766
Db	2256	CTAAGATATCCATTATAGGGCAATCGTAGGTTCAAGTATTGGTTCTGCCTATATAGCA	2315
QY	767	ctcttaatatcaaggcagctgctgtggcgctgcaggttctctggtgtgtgttcttatt	826
Db	2316	TTCTTTAAGGTTAAAGCTATTGCATTAGTAGTACTGCAGGTATACCGAGATTATATCTATT	2375
QY	827	gatgtccagatatggtcatggtt 849	
Db	2376	AGTGGTCAAAATAAATGGTTGGTT 2398	
RESULT	13		
BSTREAPR			
LOCUS	B.subtilis treA, trep and treR genes.	BCT	21-NOV-1996
DEFINITION			
ACCESSION	254245		
VERSION	254245.1	GI:1000450	
KEYWORDS	enzyme II-tre; phospho-alpha-(1,1)-glucosidase; phosphotrehalase; repressor; treA gene; trehalose operon; trep gene; treR gene.		
SOURCE	Bacillus subtilis.		
ORGANISM	Bacillus subtilis		
REFERENCE			
AUTHORS	1 (bases 1 to 4158)		
TITLE	Schock,F. and Dahl,M.K.		
	Analysis of DNA flanking the treA gene of Bacillus subtilis reveals genes encoding a putative specific enzyme II-tre and a potential regulator of the trehalose operon		
JOURNAL	Gene 175 (1-2), 59-63 (1996)		
MEDLINE	97074649		
REFERENCE			
AUTHORS	2 (bases 1 to 2543)		
TITLE	Helfert,C., Gotsche,S. and Dahl,M.K.		
	Cleavage of trehalose-phosphate in Bacillus subtilis is catalysed by a phospho-alpha-(1-1)-glucosidase encoded by the treA gene		
JOURNAL	Mol. Microbiol. 16 (1), 111-120 (1995)		
MEDLINE	95379486		
REFERENCE			
AUTHORS	3 (bases 1 to 4158)		
TITLE	Dahl,M.K.		
	Direct Submission		

JOURNAL	Submitted (27-SEP-1995) Dahl M. K., University of Erlangen-Nuernberg, Lehstuhl fuer Mikrobiologie, Staudtstr. 5, Erlangen, Fed. Rep. of Germany, 91058
FEATURES	
source	Location/Qualifiers
	1. .4158
	/organism="Bacillus subtilis"
	/strain="168 Marburg"
	/isolate="1A1"
	/db_xref="taxon:1423"
	/sex=""
	/clone_lib="genomic DNA"
	/clone="clone= pCH1"
gene	167. .1579
	/gene="trep"
CDS	167. .1579
	/gene="trep"
	/function="enzyme II tre, domains B and C"
	/note="position 1136 corresponds to position 1 of x80203"
	/citation=[1]
	/codon_start=1
	/transl_table=11
	/product="trep"
	/protein_id="CAA91014.1"
	/db_xref="GI:1000451"
	/db_xref="SWISS-PROT:P39794"
	/translation="MGELNKSARQIVAEVAGGAENTAAATHCVTRLRFALIDESKVDQE
	MLDQIDVVKGSFSTNGQFOVVGQTVNKKVYAEELVKETGIGESTDEVKKASEKNNP
	LQRAVKTADIFIPILPAIVTAGLLGNNIIITAEGLISFSTKSIQVYVPOWADLANMI
	NLIATGATFTPALIGWSAKRFGNPLGIIVGLVNLVHPDLLNANGYGAABQSGEIP
	VNLFGLVQKVGQYQVLPILLASVMLAKIEVFLT KRTPEGTQLLVAVPITLLTGF
	ASFITIGTITFAIGNVLTSLVFGSFAALGGLYGGFYSAIVTGMHTFLAVDLQ
	LIGSLGCTFLPMLALSNTAOGSAAALAMFVKKDEKOKGLSITSGISAVILGITEPAI
	FGVNLRYRPPFIIMVSSSLAGMYISSQGVASSVGVGPGIFSIQSQYWGAFAGM
	ALVLIVPFAGTYAYGRFRHK"
gene	1650. .3317
	/gene="treA"
CDS	1650. .3317
	/gene="treA"
	/standard_name="phosphotrehalase"
	/function="phospho-alpha-(1,1)-glucosidase"
	/citation=[1]
	/citation=[2]
	/codon_start=1
	/transl_table=11
	/evidence=experimental
	/product="treA"
	/protein_id="CAA91015.1"
	/db_xref="GI:1000452"
	/db_xref="SWISS-PROT:P39795"
	/translation="MKTEQTPWKKAVVYQIYKSFNDTNGVGDNGIIEKLDYLK
	TQVDVLMUTPIYDSFQHDGYDIRDYISYIPEYGTMEDFERLVSEAHKRDLLKVMDL
	VYNHTSTHKWLGDLFNROPYRDFYIMKKQENGVSPTNWSKFGGSAMELDEASGQ
	YVHLFDVTQADLNWNEEVKRVYDMHMFWEKEIDGFRDLVINLISKDRFPNAEE
	GGRSFYTDGPRVHEFLHMEKNEKVFSDYMTVGENSTVDHCIRYTPNDKELDMT
	ESFHLKYDYLTAEKWALAPDFLKLKELSDQMTGMHAGGGWALFMCNHQDPVYVS
	RYGDDCAVYKSAKMLATAIHMQGTPIYQOEEGLGNTPKFTDIDSSYRDAESLNMYH
	AFKRGMAQDDITAILQAKSRDNPVQMDATENGFTTGTTPWIPVAGNYREINAE
	ALKQDQNSFYHYKLIQIRKMDIVTEGTYEIIAKDDPNIFAYLRHGSNEKLLVINNF
	YTEAAFTLPDSLAPDEWKAEVLLTNDAREGLQNNMTVAHMSPLFIV"
gene	3352. .4068
	/gene="treR"
CDS	3352. .4068
	/gene="treR"
	/function="repressor of the trehalose operon"
	/note="position 3684 corresponds to position 2543 of x80203"
	/citation=[1]
	/codon_start=1
	/transl_table=11
	/evidence=experimental
	/product="TreR"
	/protein_id="CAA91016.1"
	/db_xref="GI:1000453"

```
/db_xref="SWISS-PROT:P39796"  
/translation="MKVNKFTIYKDIAQIEGGRWAEELTPEHELTAYQYTSRET  
VRKALHMLAQYIQIRKGSVNLNRKMGFPYSGLVSKELATGLTKETTVHKF  
GLEPSELQKQRLANDLDIWEVIRSRKDGEHVLDDKYFPRKHVPLTKELCENS  
IYBYIGELGLSISYQAQKEI VAEPTQDDBRELLDQYDHWVVRNVYFLDSTLSQY
```

BASE COUNT	1137 a	838 c	1088 g	1095 t
ORIGIN				

Query Match 13.5%; Score 205.4; DB 1; Length 4158;  
Best Local Similarity 53.1%; Pred. NO. 1.5e-44;  
Matches 485; Conservative 0; Mismatches 421; Indels 7;

Qy	1	ctcatggcatctgcgcgcttcgcgttctcttctgcgcagtgctggttggtttccacgcgcaaccaag	60
Db	650	CTGATTGCGAGCAGCGCCCTTACGTTTTTGGCTGGTGAATCGGATGGTCTGCCGTCACG	709
Qy	61	cgtttcgcgcgcaatgagttcctggccgcgcgattggtatggcgatgggtgttcocgag	120
Db	710	CGGTTTCGGGGGCAATCCGCTTCTCGGCATGTGTC-TCGGGGTATGTCGTGTCGATCCCTG	768
Qy	121	cttgggtgaacggctacgacgctggccgcaccactggctcgggcgaaatgccaaatgtgtc	180
Db	769	TTTGCTGAAATCGCTGGGGATATGGCGGCGCAGACAAAGCGGAGAGATCCCGGTATGGAA	828
Qy	181	ctcgtttgttttagatgttgccaaagcggttaccagggcaccgtgtctctctgtctggt	240
Db	829	TCATTTCGGCCCTAGAGGTGCAGAAAGTGGCTATCAGGGCCAGGTGCTCCCAATTTTGCT	888
Qy	241	ggttcttgattctggcgaacgatcgagaagttcctgcacaaaggaactcaagggaactgc	300
Db	889	TGCTTCTTATGCTGCGCAAGATTGAGGTATTTTAAACAAACGCACACCTGAGGCCAT	948
Qy	301	agacttcctgatcaactccaagtctgaagctgtgctgcaccggattccttacaatcatgc	360
Db	949	ACAGCTGCTGCTGTGCGCACCAATTAACGCTTCTCTTAACAGGATTTGGCTCTTTTATCAT	1008
Qy	361	cattggccccagcaatgcgtgggtggcgatgctgctggcacacaggtctacagggaactta	420
Db	1009	TATCGGCCCGATTACATTTGCCATCGGTAAATGACTGACTTCAGGGCTTATCTCGGTGTT	1068
Qy	421	tgaattcgggtgcgaagtcggcggtgtcgtctgttgggtctgttactaccaaatcgtcat	480
Db	1069	TGGTTCTTTTTCGGCATCTGGCGGCTCTGTTATATGCGGTTTCTACTCAGCGCTCGTGAT	1128
Qy	481	cactggtctgcaccagtccttcgcgcgaatgagctgagct-----gtttaaccaggg	534
Db	1129	TACCGGTATGTCATCACAGTTTCTTGCACTCGACCTGCAGCTCATCGGTTCAAAGCTCGG	1188
Qy	535	tggatccttcatcttcgcgaagcatctatggcttaatcgcaccagggtgcggcatgttt	594
Db	1189	CGGCACCTTTTATGGCCGATCTGGCGCTTCTTAATATCGACAAAGTTTCGGCGGCTCT	1248
Qy	595	ggcagtgcttcttcgcggaagagtgaagaactcaaggccttgcaggtgcttcagtg	654
Db	1249	GGCATTTGATGTTTTCGTGAAGGATGAGAAACAAAGGCCCTCTCCCTACATCAGGGAT	1308
Qy	655	ctcgcgttcttggtattaccggagcctgcgactctcggtgtgaaccttcgcctgcgctg	714
Db	1309	TTCTGCTTATCTTGGCATTTACAGAGCCCTGCCATCTTTGGGTTGAATCTGGGATACAGATT	1368
Qy	715	gcggtttcttcggtatcggtacgcagctatcgggtggcgcttggattgcactctttaa	774
Db	1369	CCCGTTTATCATTTGCAATGGTCAGCTCTGGCCTGGCGGAATGATATTCTTCTCAAGG	1428
Qy	775	tatcaaggcaggttgcttggcgctgcaggtttcttgggtgtgttcttctatgatgtcc	834
Db	1429	TGTGCTGGCAAGCTCGTCTGGTGTGGCGGGGTGCCGGGGATTTTCTCATCATGAGCCA	1488
Qy	835	agatatggtcattgttcttgggtgtgtgcagtttgttaaccttcttcacgtatcgctgcgcgc	894
Db	1489	GTACTTGGGCGCGTTTGCAAATCGGATGGCTATTGTATTGATCTGCTGCGGTTTTCGGAAC	1548

	/codon_start=1	
	/transl_table=11	
	/product="yfkA"	
	/protein_id="BAA23391.1"	
	/db_xref="GI:2626813"	
	/translation="MTONTKLRPTTPEFDWEAYMDVXDGMQLTNVETTTTLCNMC	
	CEHCAGVYIQLQDPNALPDIIDLLKRLBEIPRLRSISITGGEPMLSLKSVKEYVPEL	
	LKYAHGKRVQINSNTLIDIERYEWIIPYLDVLHISHNMGTVEDFAEIGFA"	
terminator	complement(2666. .2701)	
RBS	2750. .2768	
CDS	2766. .3227	
	/codon_start=1	
	/transl_table=11	
	/product="yfkB"	
	/protein_id="BAA23392.1"	
	/db_xref="GI:2626814"	
	/translation="MYPDSFASALESLKEMRQAIRLLDIDRNTWMLFGLTLPFYA	
	CSPDEDQKLLRRLRAAKNTVVRNDPGRSLRVNFDGNLIIVTDGDPPLPLNGIQTID	
	SLPAYEKWRKTEKLAKLSCHSCPVRCLGNVLVKNSIYODTFTKRQAKS"	
RBS	3217. .3230	
CDS	3228. .4070	
	/codon_start=1	
	/transl_table=11	
	/product="yfkC"	
	/protein_id="BAA23393.1"	
	/db_xref="GI:2626815"	
	/translation="MRMKETLTFEIQNKVILVAVILVGVFIINRLVOLFEKFRTD	
	PIEBKKETTESLVRSYQTYATYTGTFYFVLSILFVHDGKILAGAGVAGVIGFGAQS	
	LKQVLQAGVLEVRQHLKGDYTVNNLVNTGVEEIGLSLQIRESGKGLITTSNVEY	
	ROIENYINDFWRITESLISFKEDPDRVYSVLEEACDMLNEELRDSLKRDENFNTEP	
	FOIHGIALNKINRGVEFTVKGVMVKDDDFYSASLAVRRVLVRLQYNNQVMLEEAVERI	
terminator	ERTQ"	
	complement(4074. .4106)	
CDS	complement(4109. .4903)	
	/codon_start=1	
	/transl_table=11	
	/product="yfkD"	
	/protein_id="BAA23394.1"	
	/db_xref="GI:2626816"	
	/translation="MMKKLFHSTLIVLFFSFQVQPIHAKOKFPKVPNSVASISKENT	
	YPNASQDPHLPQSLKAKELLDSHVEKVIENPHLILKMLNESNISGTPPLAVGYRATIFLG	
	KWALIGYESNETVANEWKYKILNTNRADNRGGKTAEMHYAQEQYRVKGGVITAKVPNAE	
	DVKSMMQKMKRKNLPLAFETVIGAGTKRDQYIKVAPKKIYGLYHAYAPAVNKEKGVT	
	YGEVLYLKGKRRLLVKNVTSSQIGAWIPVQDHVTFGFQLSLPR"	
RBS	complement(4901. .4913)	
CDS	complement(4974. .6029)	
	/codon_start=1	
	/transl_table=11	
	/product="yfkE"	
	/protein_id="BAA23395.1"	
	/db_xref="GI:2626817"	
	/translation="MNSRIFELVAGVPLSVIGSLMHPASVLFAVYCVTIILASLYM	
	GRATESIINGPRIGGLKNATGVNAGVELLISLPALKEGTGIVLASLGTSGVIGNLILL	
	VAGUSFVGGLKRYKRFENHDAHNSNGLLIFAIIVAFVPEVFSVGMGNASKNLISI	
	GISIMILLVYAAVLQHLTHRGYQPNNAQTEEEEPWSGKVATIVLFAATIVVA	
	YISNLVHTFSVAEQFQWSLFGTIVTIVAGNAEHAASAIIMAFKNKMDIAVEIAY	
	GSTLQIAMFVAPVLVICSIFEPTSMPLVFTLPVLVAMVSALLMIAISNDGDSNWFEG	
	ATLAAVINAIGFELL"	
	complement(6027. .6046)	
RBS	6180. .6199	
RBS	6197. .7372	
CDS		
	/codon_start=1	
	/transl_table=11	
	/product="yfkF"	
	/protein_id="BAA23396.1"	
	/db_xref="GI:2626818"	
	/translation="MSRHFHFLVLLVSIQSGQLMLPVISIIFETNGESAAINGHLH	
	ATGLYIGVLASPMPEAPLRKLGKPLIVMGSGIVLSLFGFTWLOSVWVFLRLFI	
	GIGHMLHFSOTWVTSKNSQNGRNLISYIGLSFGIGFAAGPMPYLKLSLSLPI	
	VSGCISLPALVFFELQYAPETSPHETKSDNSFRFYPQAMLFQWAFVMPFGIGLE	
	TALNGSIFALRGIGISDAVAILPAFASIIQFPLGILSDKYGRRNVLLVILT	
	GALCFITAGVPSPYVGGGFTAGMAVGSTFTIGISTMTDILLPPHLLPAGNLICGT	
	FSLGSIILGPMGVNMTFFESANLIFYETITLSSVNLALVLGPKKSPAPETVSSSS"	





```
/product="yfkx"
/protein_id="BAA22292.1"
/db_xref="GI:2443221"
/translation="MISYIVOTLIVCIAIYAYENKFRSANNLTKWAFSLLIAGSAFL
WIYMRVNLPLRLGLHFKYIPF"
RBS 4563. 4579
CDS 4577. 5653
/codon_start=1
/transl_table=11
/product="yfkx"
/protein_id="BAA22293.1"
/db_xref="GI:2443222"
/translation="MDKTSAYOGFFGALYTLAYGLKHAPILMIESAKONAWHSYILG
VWIVPALMLHRLMKHODKNYIELLSDSSPIAGRIIILLESYFLLINAHDIRFFI
NLINLELPTPMALVGVIIFVAICAREKETLRMAQIFLFPGLVLFPLPFTLA
TQELQNLTPVEGLIPLQSYAFAGTMGELIILPLLSNRSPKLYITFAILLGAL
LLVAMPSSISVFNSTSTFFDPAYAVIRQIRITDLDRSDLIIAFAFPIVIMVKA
GSLVIVYGSLHSLKIDPKAMYPPTGMSVVCWFELNTNQLIDFNRIKPIINVI
SLLPLLIYLIKSKALFGAKAH"
CDS complement(5650. 7065)
/codon_start=1
/transl_table=11
/product="yfkx"
/protein_id="BAA22294.1"
/db_xref="GI:2443223"
/translation="MERLWIEHISDMLGPPILILLTGLYFTILLKGFQRYPL
YIFQTTGSGKKPKGEGTVPLOALSALSSITGAANIVGPAAIMFGGGAFFWMW
LIAFLFAIRKESVLAHYREKNEQGEYVGGPMYITKGLRMKWLGVFFSVALIVEL
IPSIWQGVNSVSLAETFFSPNKIYAGIGIAFLIGLVVIGGVRIGKVFEPVPLMAG
AYAGALLIYLMNLSVPAPFSLVSNFNAFTSSAVGGFAGAAEAETVWGFARGLYSN
EAGGTAPIAHAAMTDPHVRQGFWSVIGVIDTILICTTTAFIVLASGVTWTKNASN
DPAALTTAAFOHYFGSGSGYFVSLSVFEFVSTIMVIFYGVKQAELEFGLAGHVIK
FYVLAIIIGAGAKAIWGLVDLALFIVVPNVIALLLLSRKKYKALYTEFTSEQYY
LKDIRTKQKPYTPTEKAKNS"
RBS complement(7063. 7077)
terminator complement(7149. 7175)
CDS complement(7190. 7450)
/codon_start=1
/transl_table=11
/product="yfkx"
/protein_id="BAA22295.1"
/db_xref="GI:2443224"
/translation="MGYRIKLNAGMAEDGALKKELMTIDHLEKCYDEKEREQLIRK
KSEKQLLDKLVDRKGMFSKPSAFYKNKVYDLGRNRPSS"
CDS complement(7432. 7584)
/codon_start=1
/transl_table=11
/product="yfkx"
/protein_id="BAA22296.1"
/db_xref="GI:2443225"
/translation="MMDFSHIVSEDKIKRAIKDGFQNLPGMKRPLPKMTRHTCNRC
AWAIVF"
RBS complement(7448. 7466)
RBS complement(7582. 7596)
CDS complement(7600. 7737)
/codon_start=1
/transl_table=11
/product="yfkx"
/protein_id="BAA22297.1"
/db_xref="GI:2443226"
/translation="MREKVAKNAVESTFREDITCKTRYLSRNKGIKWYIENCMIKYK
V"
RBS complement(7735. 7752)
RBS 7824. 7840
CDS 7838. 7987
/codon_start=1
/transl_table=11
/product="yfkx"
/protein_id="BAA22298.1"
/db_xref="GI:2443227"
/translation="MKTFIKRGIAFFLIAVLLWIKTYGVVLFNFGIDNTIQKIL
LFVNPLSSSLFPLGFLGLKKLOOTAIIVIFHLSFLLYANIVYRFPNDFITPIVI
MQAKTNGQLGDSAFSLMRPTDAFYFIDTIIILAIKVNKPAETSSSKSFRIIFASS
```

```
Db 626 TATCGGCCCGATTACATTTCGGATCGGTAATGTACTGACTCAGGGCTTATCTCGGTGTT 567
QY 421 tgattcgggtggtccagtcggtggtctgctcttctggtctggtctactcaaccaatcgtcat 480
Db 566 TGGTCTTTTGGCCGACTGGGCGTCTGTATATGGCGGTTTCTACTCAGGCTCGTGAT 507
QY 481 cactggtctgcaccagtccttcccgcaattgagctgagact-----gtttaaccagg 534
Db 506 TACCGGTATGCATCACACGTTTCTTGCAGTCGACCTGCAGCTATCGGTTCAAAGCTCGG 447
QY 535 tgatccttcctcctgcgaagcgcacatctatggtctaataatgcgccagggtgcggcatgtt 594
Db 446 CGGCACCTTTTATGGCCGATGCTGGCGCTTCTTAATATCGCACAGGTTGGCGGCTCT 387
QY 595 ggcagtgcttcttcctggtgcgaagagtgaaagctcaaggccttgccagggtgcttcaggtgt 654
Db 386 GGCCATGATGTTTATCGTGAAGGATGAGAAACAAAAGGCCCTCTCCCTCACATCAGGGAT 327
QY 655 ctccgctgttcttggtattacgagcctgcgatcttctggtgtgaaccttcgacctgcgtg 714
Db 326 TTCTGCTTATCTTGGCATTACAGAGCCTGCCATCTTTGGGGTGAATCTGGGATACAGATT 267
QY 715 gccgttcttcacgtgtatcggtaccgcagctatcggtggtggcgtttgattgcactcttta 774
Db 266 CCCGTTTATCATTCGAATGTCTAGCTCTGGCCTGGCGGAAATGTATATTTCTTCTCAAGG 207
QY 775 tatcaagcgagttgcgttggggcgtgcaggtttcttgggtgtgttcttctattgatgtcc 834
Db 206 TGTGCTGGCAAGCTCCGTCGGTGTGCGGCGGTGCCGGGATTTTCTCAATCATGAGCCA 147
QY 835 agataggtcatgttcttgggtgtgtgcagttgttaccttcttcacgtcattcggcgagc 894
Db 146 GTACTGGGGCGCGTTTGCATCGGATGGGCTATTGTATTGATCGTGCCTTTGCCGGAAC 87
QY 895 gattgettatg 905
Db 86 ATACGGGTATG 76
```

Search completed: March 22, 2002, 08:49:30  
Job time: 7679 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 21, 2002, 16:21:14 ; Search time 25.85 Seconds  
(without alignments)  
1379.098 Million cell updates/sec

Title: US-09-604-231-2  
Perfect score: 2363  
Sequence: 1 MAMVFPVLVNGYDVAAATMAA.....IEAGANLLNVAKKEAVPATP 468  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues  
Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_68:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1058.5	44.8	651	2 S44257	phosphotransferase
2	953.5	40.4	664	1 B32243	phosphotransferase
3	701	29.7	480	2 S39978	scrA protein - Sta
4	695.5	29.4	630	2 H83686	PTS system, beta-g
5	673.5	28.5	636	2 C83724	PTS system, beta-g
6	673	28.5	632	2 S68599	phosphotransferase
7	659	27.9	479	2 F82432	PTS system, sucros
8	657	27.8	609	2 I40406	beta-glucoside per
9	646	27.3	479	2 JQ0781	sucrose uptake pro
10	645	27.3	609	2 T47097	hypothetical prote
11	640	27.1	631	2 B42603	beta-glucoside-spe
12	588.5	24.9	636	2 D86807	hypothetical prote
13	570	24.1	470	2 C69725	phosphotransferase
14	569.5	24.1	470	2 H83926	PTS system, trehal
15	567.5	24.0	625	2 C25937	phosphotransferase
16	489	20.7	456	2 S62331	phosphotransferase
17	487	20.6	458	2 H83881	PTS system, sucros
18	473	20.0	455	1 W0EBST	phosphotransferase
19	467	19.8	372	2 I39868	sac operon regulat
20	466	17.2	459	2 JU0293	levansucrase synth
21	390	16.5	460	2 A39938	phosphotransferase
22	367	15.5	473	2 C65236	phosphotransferase
23	360	15.2	473	2 A86122	PTS system enzyme
24	342.5	14.5	453	2 F84096	PTS system, sucros
25	341.5	14.5	692	1 S46953	phosphotransferase
26	339	14.3	478	2 B82263	PTS system, trehal
27	321.5	13.6	699	1 W0BSGS	phosphotransferase
28	317.5	13.4	651	1 S18607	phosphotransferase
29	314	13.3	195	2 D86766	hypothetical prote

30	305.5	12.9	675	2 D83755	PTS system, glucos
31	298.5	12.6	648	1 W0EC2N	phosphotransferase
32	297.5	12.6	648	2 E85567	hypothetical prote
33	289.5	12.3	675	1 S46952	phosphotransferase
34	287.5	12.2	631	2 D69750	phosphotransferase
35	285	12.1	521	2 C86678	hypothetical prote
36	283.5	12.0	446	2 B69744	sucrose phosphotra
37	281	11.9	454	2 C86766	hypothetical prote
38	271.5	11.5	173	2 C83839	PTS system, glucos
39	270.5	11.4	485	2 G65051	phosphotransferase
40	265.5	11.2	485	2 B85920	hypothetical prote
41	256	10.8	189	2 F70169	phosphotransferase
42	255	10.8	168	2 H69940	phosphotransferase
43	254.5	10.8	634	2 A32241	lactose transport
44	253.5	10.7	474	2 G85885	probable PTS enzym
45	250.5	10.6	474	2 D65017	hypothetical prote

ALIGNMENTS

RESULT  
S44257 1

phosphotransferase system enzyme II (EC 2.7.1.69), sucrose specific - Pediococcus pen  
C:Species: Pediococcus pentosaceus  
C:Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 16-Jul-1999  
C:Accession: S44257  
R:Leenhouts, K.K.J.; Bolhuis, A.A.; Kok, J.J.; Venema, G.G.  
A:Description: The sucrose and raffinose operons of Pediococcus pentosaceus PPE1.0.  
A:Reference number: S44252  
A:Accession: S44257  
A:Molecule type: DNA  
A:Residues: 1-651 <LEE>  
A:Cross-references: EMBL:232771; NID:g493728; PIDN:CAA83668.1; PID:g475968  
C:Genetics:  
A:Gene: scrA  
C:Superfamily: phosphotransferase system enzyme II sucrose-specific; phosphotransfera  
C:Keywords: phosphotransferase  
F:488-640/Domain: phosphotransferase system glucose-specific enzyme II, factor III ho





C;Accession: F82432  
R;Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;  
Chardonson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, B.  
L, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.  
Nature 406, 477-483, 2000  
A>Title: DNA Sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*.  
A;Accession: F82432  
A;Reference number: A82035; MUID:20406833  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-479 <HEI>  
A;Cross-references: GB:AE004395; GB:AE003853; NID:g9658068; PIDN:AAF96554.1; GSPDB:GN001  
A;Experimental source: serogroup O1; strain N16961; biotype El Tor  
C;Genetics:  
A;Gene: VCA0653  
A;Map position: 2  
C;Superfamily: phosphotransferase system sucrose-specific enzyme II, factor II

Query Match 27.9%; Score 659; DB 2; Length 479;  
Best Local Similarity 49.3%; Pred. No. 3.7e-38;  
Matches 133; Conservative 46; Mismatches 87; Indels 4; Gaps 1;

QY 1 MAMVPSLVNGDYVAATMAAGMPMWSLGLDVAAGYQGTVPVLVVSWMILATIEKFLH 60  
Db 194 MLMVHPOLLNCGFGSASVSGTPTWNILGFIEKVGYSGLPVLVSVILAKIENGLR 253  
QY 61 KRLKGTADFLITPVLTLTLLLTFTTFAIGPAMRWGVDLAHGLOGLYDFGPGVGLLFL 120  
Db 254 KIVPSVDNLTLMPLAIFITGELTFTVVGPLTRDVGMLNWLVDYSAGFVGALFGF 313  
QY 121 VYSPVITVLTGHQSPFPPIELELF-----NOGGSFIFATASMANIAQGAACLAFFFLAKSEKL 176  
Db 314 IYAPFVITGMHSHFIAETOLLADIVTGTFFIPPAAMSNIAQGAALAVGVTKETKL 373  
QY 177 KGLAGAGSVAVLGITPAIFGNVLRWPFIFIGTAAIGGALIALFNKIKAVAGAGF 236  
Db 374 KGVAIPSGVTALLGITEPAMFGVGNLRLPYFIAAICGAALASAFITLNFVKAQALGAAGL 433  
QY 237 LGVVISIDAPDMVFLVCAVVTFTFAAGAAI 266  
Db 434 PGIISINPOQGIYIMGMALISFAAFALT 463

RESULT 8  
I40406  
beta-glucoside permease - *Bacillus subtilis*  
C;Species: *Bacillus subtilis*  
C;Date: 12-Aug-1996 #sequence\_revision 12-Aug-1996 #text\_change 20-Jun-2000  
R;Le Coq, D.; Lindner, C.; Kruger, S.; Steinmetz, M.; Stulke, J.  
J. Bacteriol. 177, 1527-1535, 1995  
A>Title: New beta-glucoside (bgl) genes in *Bacillus subtilis*: the bglp gene product has  
A;Reference number: I40406; MUID:95189730  
A;Accession: I40406  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-609 <RES>  
A;Cross-references: EMBL:X85408; NID:g1037169; PIDN:CAA59697.1; PID:g1037172  
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berte  
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chd  
A.; Ehrlich, S.D.; Emmerson, P.F.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.  
Nature 390, 249-256, 1997

A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal  
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.  
Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino  
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau  
Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete  
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanl  
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se  
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiya  
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida  
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.  
A>Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*  
A;Reference number: A69580; MUID:98044033  
A;Accession: A69594  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-287, F', 289-433, G', 435, S', 437-448, HR', 451-548, M', 550-551, S', 553-60  
A;Cross-references: GB:299124; GB:AL009126; NID:g2636442; PIDN:CAB15963.1; PID:g26364  
A;Experimental source: strain 168  
C;Genetics:  
A;Gene: bglp  
C;Superfamily: phosphotransferase system enzyme II sucrose-specific; phosphotransfera  
F;458-609/Domain: phosphotransferase system glucose-specific enzyme II, factor III ho

Query Match 27.8%; Score 657; DB 2; Length 609;  
Best Local Similarity 33.4%; Pred. No. 6.7e-38;  
Matches 153; Conservative 84; Mismatches 181; Indels 40; Gaps 8;

QY 13 DVAATMAAGMPMWSLGLDVAAGYQGTVPVLVVSWMILATIEKFLHKLKGFADFLIT 72  
Db 186 DLTAALLGAGK-PI-SFICLPVTAATYSTVTPILLISWIASYVEKWDRETHASLKLIIV 243  
QY 73 PVITLLITGLTFTTATGPMRWGVDLAHGLOGLYDFGPGVGLLFLGVSPVITGLHQ 132  
Db 244 PTFLLIVVPLTLITVGLGAILGEYLSSGVNVLFDHAGLVAMILLAGTSLIIMTGMHY 303  
QY 133 SFPIELELENQGG-SFIFATASMANIAQGAACLAFFFLAKSEKLKLAGASGVSAVLGI 191  
Db 304 AFVPIMINNTAONGHDYLLPAMFLANNGAGASFAVFLSRNKKFKSLALTTSITMGI 363  
QY 192 TEPAIFGVNLRWPFIFIGTAAIGGALIALFNKIKAVAGCA-AGFIVGVVISIDAPDMVF 250  
Db 364 TEPAITGVNMLKPKFAAALIGGAAGAFYGMTGVASYIVGNGAGLSPVIFIGPTFIYA 423  
QY 251 LVCAVVTFTFAAGAAITAGLYLVRNGSIDPDATAAPVPAGTTKAAEAFAEFSNDSTII 310  
Db 424 MIGLVTAFAAETAAAYLLGFEDVPSDGS-----QQPAVHEGSRRII 464  
QY 311 QAPLTGEATIALSSVSDAMFASGKLGSGVAIVPTKQLVSPVSGKIVVAFSPGHAFVATK 370  
Db 465 HSPKGEVKALSEVKDGVFSAGVMGKGFATEPEEGEVVSPVSGSVTTFIKTKHAIGTS- 523  
QY 371 AEDGSNVDIIMHIGFTDVNUNGTHFNPLKQKQDEVKAGELLCFDDIDAKAAGYEVVTP 430  
Db 524 ---DQGAELIHLGLDVTKLEGMFTAHIEKQKVAEPDPLVSDFLRQIKAAAGYDVITPV 580  
QY 431 VVSNYKK--TGPVNTVGLGEIEAGANLLNVAKKEAVPA 466  
Db 581 IVTNDQYSFSPVKEIG-----KVQPREALLA 607

RESULT 9  
JQ0781  
sucrose uptake protein - *Vibrio alginolyticus*  
N;Alternate names: enzyme II-sucrose protein  
C;Species: *Vibrio alginolyticus*  
C;Date: 04-Sep-1998 #sequence\_revision 04-Sep-1998 #text\_change 18-Jun-1999  
C;Accession: JQ0781  
R;Blatch, G.L.; Scholle, R.R.; Woods, D.R.  
Gene 95, 17-23, 1990  
A>Title: Nucleotide sequence and analysis of the *Vibrio alginolyticus* sucrose uptake-  
A;Reference number: JQ0781; MUID:91071601  
A;Accession: JQ0781

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-479 <BLA>

A:Cross-references: GB:M76768; GB:M30194; NID:g155261; PIDN:AAA27555.1; PID:g155262

C:Genetics:

C:Gene: scrA

C:Superfamily: phosphotransferase system sucrose-specific enzyme II, factor II

Query Match 27.3%; Score 646; DB 2; Length 479;

Best Local Similarity 46.5%; Pred. No. 2.9e-37;

Matches 128; Conservative 54; Mismatches 87; Indels 6; Gaps 3;

QY 1 MAWPELVNGYDVAATMAAGEMPWLSFLGLDVAQAGYQGTVPVLVYVSWILATIEKFLH 60

DB 194 MLWVHPDLLNGWFGGASVSGNIPWNILGFEIQKVGQSVLPVLVSAFILAELVELGLR 253

QY 61 KRLKGTADPLTPVLTLLTGLTFLATGAPAMRWGVDVLAHGLQGLYDFGPGVGLLFLG 120

DB 254 KVIPSLDNLTPLLAIFTAGLLTFTVVGPTTRDIFGLDGLNWLNTAGFVGAVFGL 313

QY 121 VYSPVITGLHQSFPPIELELF-----NOGGSFIFATASMANIAQAACLAIVFLAKSEKL 176

DB 314 IYAPFVITGMHSHFATETQALLADIATTGCTEFPFAAMSNVSGAAALAVGMSKDKKM 373

QY 177 KCLAGASGVAVLGIETAPFCVNLRLRPFFIGICTAAGGALLALFNKAVAGAAAGF 236

DB 374 KGIAIPSGVTGLGITEPAMFGVNLKLRYPFFIAACAAALSSAFITMFMVAKAALGAAGL 433

QY 237 LGWVSIDAPDMVFLVCA-VVFFTAFAAGAAIAYGL 270

DB 434 PGIISI-TPDKIGYIAGVIAFLTAFLVTLVGI 467

RESULT 10

T47097

hypothetical protein N17C [imported] - Bacillus subtilis

C:Species: Bacillus subtilis

C:Date: 17-Mar-2000 #sequence\_revision 17-Mar-2000 #text\_change 21-Jul-2000

C:Accession: T47097

R:Yoshida, K.; Sano, H.; Seki, S.; Oda, M.; Fujimura, M.; Fujita, Y.

Microbiology 141, 337-343, 1995

A:Title: Cloning and sequencing of a 29 kb region of the Bacillus subtilis genome contain

A:Reference number: 224350; MUID:95219088

A:Accession: T47097

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-609 <YOS>

A:Cross-references: EMBL:D31856; NID:g603765; PIDN:BAA06652.1; PID:g603778

A:Experimental source: strain BGSC1A1

C:Genetics:

C:Gene: N17C

C:Superfamily: phosphotransferase system enzyme II sucrose-specific; phosphotransferase

Query Match

Best Local Similarity 27.3%; Score 645; DB 2; Length 609;

Matches 154; Conservative 86; Mismatches 178; Indels 40; Gaps 11;

QY 13 DVAATMAAGEMPWLSFLGLDVAQAGYQGTVPVLVYVSWILATIEKFLHKLKGTADFLIT 72

DB 186 DLTALLGAGK-PI-SFGLPVTAAVTSYVIFILLISWIASVVERWIDRTHASLKLIVV 243

QY 73 PVLTLLTGLTFLATGAPAMRWGVDVLAHGLQGLYDFGPGVGLLFLVSPVITGLHQ 132

DB 244 PFTLLIVVPLTLITVGLAIGELVSSVNYLPDHAGLVAMIFLAGTFLSLIIMTGMHY 303

QY 133 SPPIELELFNQG-SFIFATASMANIAQAACLAIVFLAKSEKLKGLAGASGVAVLGI 191

DB 304 AFVPTMINIAQNGHDIYLLPAMFGVNLKLRYPFFIAACAAALSSAFITMFMVAKAALGAAGL 363

QY 192 TPAIFGVNLRLRPFFIGICTAAGGALLALFNKAVAGAAAGF 250

DB 364 TEPAMYGVNMLKPKFAAALIGGAAGGAFYGMTGVASYIVGNGAGLPSIPVFIGPTFIYA 423

QY 251 LVCVVTFPIAAGAAIAYGLYLVRNNGSIDPDATAPVPAGTTKAEAEAPAFESNDSTII 310

DB 424 MIGLVIAF--AAGTSAAYLL-----GFED-----VPS-----HRSQQQPAVHEGSRRII 464

QY 311 OAPLGEATALSVDAMFASGKLGSGVAIVPTKGOLVSPVSGKIVAVPSPGSHAFAPVRK 370

DB 465 HSPITGEVKALSEVDKGVFSAGVMKGFAPFEEBEVSPVRGSVTTIFKTHTAIGITS- 523

QY 371 AEDGSNDILMHIGFTVNLNGTHFNPLKKQDGEVKAAGLCEFFDIDAITKAAGYEVTTPI 430

DB 524 ---DQGAELIHLGLTVKLEGWFTAHMKESDKVAPGDPVLSFDLEQIKAAAGYDVTVP 580

QY 431 VYSNKK--TGPVNTYGLGEIEAGANLLNVAKKEAVPA 466

DB 581 IVTNTDQYSFSPVKEIG-----KVQPKALLA 607

RESULT 11

B42603

beta-glucoside-specific phosphotransferase system-dependent permease - Erwinia chrysae

C:Species: Erwinia chrysanthemi

C:Date: 04-Sep-1998 #sequence\_revision 04-Sep-1998 #text\_change 16-Jul-1999

C:Accession: B42603

R:el Hassouni, M.; Henrissat, B.; Chippaux, M.; Barras, F.

J. Bacteriol. 174, 765-777, 1992

A:Title: Nucleotide sequences of the arb genes, which control beta-glucoside utilization

olase family including enzymes from eubacteria, archaeobacteria, and humans.

A:Reference number: A42603; MUID:92121114

A:Accession: B42603

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-631 <ELI>

A:Cross-references: GB:M81772; NID:g148385; PIDN:AAA24814.1; PID:g148387

A:Note: sequence extracted from NCBI backbone (NCBIN:77535, NCBIP:77537)

C:Genetics:

C:Gene: arbF

C:Superfamily: phosphotransferase system enzyme II sucrose-specific; phosphotransferase

F:479-631/Domain: phosphotransferase system glucose-specific enzyme II, factor III ho

Query Match 27.1%; Score 640; DB 2; Length 631;

Best Local Similarity 33.8%; Pred. No. 1e-36;

Matches 154; Conservative 86; Mismatches 200; Indels 16; Gaps 8;

QY 3 MYFPSLVNGYDVAATMAAGEMPWLSFLGLDVAQAGYQGTVPVLVYVSWILATIEKFLHKL 62

DB 184 LVHPSMIAAFN--AMQAPDHSTLHFLGIPITFINYSSSVIPILFASWVSKLEKPLNRW 240

QY 63 LKGTADFLTPVLTLLTGLTFLATGAPAMRWGVDVLAHGLQGLYDFGPGVGLLFLVY 122

DB 241 LHANIRNFTPLLCIVISVPLTFLGIPSTWLSQMLAGGYQWLYGLNLSLLAGAVNGALW 300

QY 123 SPIVTGLHQSFPPIELELFNQ--GGSFIFATASMANIAQAACLAIVFLAKSEKLKGLAG 181

DB 301 QVCVIFGLHWGFVPLMLNFSVIGHDTLLPLVPVAVGGAGATGLVLLRTQDLKRKGIAG 360

QY 182 ASGSVAVLGTEPPIAIFGVNLRWRPFFIGITAAIGGALLALFNKAVAGAAAGFGLGVYS 241

DB 361 SAFAAIFGITEPVAVGVTLPLRRPFIICGIGCALGAAMVGAHTMTYSFGPPSIFSTQ 420

QY 242 IDAPDMVFLV-CAVVTFTIAFG-AAIAYGLYLVRNNGSIDPDATAPVPAGTTKAEAE 299

DB 421 VTPPTGVDSVNAAVIGTLLAFATAALTSWSFGVPKD---ETQPAADSPA--VLAETQA 475

QY 300 PAFESNDSTIIQAPLTGEAIALSSVSDAMFASGKLGSGVAIVPTKGOLVSPVSGKIVAV 359

DB 476 NAGAVDETLF-SPLAGEVLLLEQVADRTFASGVMKGIATIRPTQGLRXPVDTVASUF 534

QY 360 PSCHAFVTRKAEDEGSDNVDILMHIGFTVNLNGTHFNPLKKQDGEVKAAGLCEFFDIDA 419

DB 535 KTHHAIGLASR----GGAELIHVGIDTVLRDGRYTFPHRVGVDVVRQDGLLEFDGPAI 590







**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 21, 2002, 16:21:59 ; Search time 29.76 seconds

(without alignments)  
1062.320 Million cell updates/sec

Title: US-09-604-231-2

Perfect score: 2363

Sequence: 1 MAMVFPVLVNGYDVAAATMAA.....IEAGANLLNVAKKEAVPATP 468

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 322656 seqs, 67552660 residues

Total number of hits satisfying chosen parameters: 322656

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending Patents\_AA\_New.\*

- 1: /cgn2\_6/ptodata/1/paa/PCT\_NEW\_COMB.pep.\*
- 2: /cgn2\_6/ptodata/1/paa/US05\_NEW\_COMB.pep.\*
- 3: /cgn2\_6/ptodata/1/paa/US07\_NEW\_COMB.pep.\*
- 4: /cgn2\_6/ptodata/1/paa/US08\_NEW\_COMB.pep.\*
- 5: /cgn2\_6/ptodata/1/paa/US09\_NEW\_COMB.pep.\*
- 6: /cgn2\_6/ptodata/1/paa/US09\_NEW\_COMB.pep.\*
- 7: /cgn2\_6/ptodata/1/paa/US10\_NEW\_COMB.pep.\*
- 8: /cgn2\_6/ptodata/1/paa/US60\_NEW\_COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	660	27.9	627	1 PCT-US02-03987-13467	Sequence 13467, A
2	660	27.9	627	7 US-09-815-242-13467	Sequence 13467, A
3	660	27.9	627	7 US-10-072-851-13467	Sequence 13467, A
4	396	16.8	473	6 US-09-897-516-6783	Sequence 6783, Ap
5	345.5	14.6	679	1 PCT-US02-03987-5658	Sequence 5658, Ap
6	345.5	14.6	679	6 US-09-815-242-5658	Sequence 5658, Ap
7	345.5	14.6	679	6 US-10-072-851-5658	Sequence 5658, Ap
8	345.5	14.6	681	1 PCT-US02-03987-12270	Sequence 12270, A
9	345.5	14.6	681	6 US-09-815-242-12270	Sequence 12270, A
10	345.5	14.6	681	7 US-10-072-851-12270	Sequence 12270, A
11	338.5	14.3	484	1 PCT-US02-03987-10809	Sequence 10809, A
12	338.5	14.3	484	6 US-09-815-242-10809	Sequence 10809, A
13	338.5	14.3	484	6 US-10-072-851-10809	Sequence 10809, A
14	334	14.1	473	6 US-09-897-516-6540	Sequence 6540, Ap
15	323	13.7	484	1 PCT-US02-03987-12272	Sequence 12272, A
16	323	13.7	484	6 US-09-815-242-12272	Sequence 12272, A
17	323	13.7	484	7 US-10-072-851-12272	Sequence 12272, A
18	305.5	12.9	455	1 PCT-US02-03987-13794	Sequence 13794, A
19	305.5	12.9	455	6 US-09-815-242-13794	Sequence 13794, A
20	305.5	12.9	455	7 US-10-072-851-13794	Sequence 13794, A
21	300	12.7	687	1 PCT-US02-03987-5822	Sequence 5822, Ap
22	300	12.7	687	6 US-09-815-242-5822	Sequence 5822, Ap
23	300	12.7	687	7 US-10-072-851-5822	Sequence 5822, Ap
24	300	12.7	719	1 PCT-US02-03987-12942	Sequence 12942, A
25	300	12.7	719	6 US-09-815-242-12942	Sequence 12942, A

26	300	12.7	719	7 US-10-072-851-12942	Sequence 12942, A
27	298.5	12.6	648	1 PCT-US02-03987-10089	Sequence 10089, A
28	298.5	12.6	648	6 US-09-815-242-10089	Sequence 10089, A
29	298.5	12.6	648	7 US-10-072-851-10089	Sequence 10089, A
30	292	12.4	551	1 PCT-US02-03987-10500	Sequence 10500, A
31	292	12.4	551	6 US-09-815-242-10500	Sequence 10500, A
32	292	12.4	551	7 US-10-072-851-10500	Sequence 10500, A
33	288.5	12.2	439	1 PCT-US02-03987-5660	Sequence 5660, Ap
34	288.5	12.2	439	6 US-09-815-242-5660	Sequence 5660, Ap
35	288.5	12.2	439	7 US-10-072-851-5660	Sequence 5660, Ap
36	250.5	10.6	474	1 PCT-US02-03987-10243	Sequence 10243, A
37	250.5	10.6	474	6 US-09-815-242-10243	Sequence 10243, A
38	250.5	10.6	474	7 US-10-072-851-10243	Sequence 10243, A
39	247.5	10.5	169	6 US-09-897-516-6596	Sequence 6596, Ap
40	233.5	9.9	482	6 US-09-897-516-7609	Sequence 7609, Ap
41	229	9.7	280	1 PCT-US02-03987-12479	Sequence 12479, A
42	229	9.7	280	6 US-09-815-242-12479	Sequence 12479, A
43	229	9.7	280	7 US-10-072-851-12479	Sequence 12479, A
44	226.5	9.6	842	1 PCT-US02-03987-11950	Sequence 11950, A
45	226.5	9.6	842	6 US-09-815-242-11950	Sequence 11950, A

#### ALIGNMENTS

RESULT 1  
PCT-US02-03987-13467  
; Sequence 13467, Application PC/TUS0203987  
; GENERAL INFORMATION:  
; APPLICANT: Elittra Pharmaceuticals, Inc.  
; TITLE OF INVENTION: Methods for Identifying the Target of a Compound which Inhibit  
; TITLE OF INVENTION: Proliferation  
; FILE REFERENCE: ELITRA.028VPC  
; CURRENT APPLICATION NUMBER: PCT/US02/03987  
; CURRENT FILING DATE: 2002-02-02  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09  
; NUMBER OF SEQ ID NOS: 15811  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13467  
; LENGTH: 627  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
PCT-US02-03987-13467

Query Match 27.9%; Score 660; DB 1; Length 627;

Best Local Similarity 34.3%; Pred. No. 8.1e-41;

Matches 159; Conservative 92; Mismatches 187; Indels 26; Gaps 10;

QY	1	MAMVFPVLVNGYDVAAATMAAGEMPMSLFGDVAQAQYQGTVPVLVSVILATIEKFLH	60
DB	182	MLLVSGSLFNAMAVA---QGGEVTAMNFGF-IPVVGLOGSVLPFAFIIGVVGAKFEKAVR	237
QY	61	KRLKGADFLITPVLTLTLTGLFTTAIGPAMRWGVDLAHLQGLVDFGPGVGGLLFGL	120
DB	238	KVDPVDILLVTPFVILLVMSILGLFVGPVHVVENYILIAIKAILSMFGLGGFLIG	297
QY	121	VYSPVITGLHQSPFPIELFNOGGSFTF-ATASMANIAQAACLAFFLAKSEKLKL	179
DB	298	VHQLIVVSGVHHFNLEVLQAAADHANPENAIITAAMTAQAATVAVGVKTKPKLKT	357
QY	180	AGASGVSVNLGITEPAIFGVNLRWPFPGTIGTAIGGALIALFNKAKALGAAGFLG-	238
DB	358	APPAALSALGITEPAIFGVNLRWPFPGTIGTAIGGALIALFNKAKALGAAGFLG-	417
QY	239	VVSIDAPDMVFMVCAVVTFFIAFGAIAVGLVLRNRSIDDPATAAPVPACTTKAAE	298
DB	418	MLTVGNGQFQYULMVAVSFALGFALTYMEGY-----EDEVDATAAKQAEVKEE	470
QY	299	-APAEFSNDSTIIQAPLTGEALISVSDAMFASGLGSGVAIVPTKQLVSPVSGKIV	357
DB	471	VAPAAQLQNETLV--TPIVGDVVALADVNDPVFSSGANGQGIKPSGVVYALADAEVSI	528



Db 471 VAPALQNETLV--TPIVGVVALADVNDPVSSAMGGGIAVKSGQVVALADAEVSI 528  
Qy 358 APPSGHAFVTRKABDGNVDILMHIGFDTVNLNGTHFNPLKKGDEYKAGELLCEFDID 417  
Db 529 APPTGHAFGLKTR----NGAEVLHVGIDTVSMNGDGFPEAKVQAQGNKVKAGDVLGTGFDN 584  
Qy 418 AIKAAGYEVTPPIVSN---YKKTGPVNTYGLGEIEAGANLLNV 458  
Db 585 KIAAAGLDDTTMTVITNTADYASVAPVAT---GSVSKGDVIEV 625

RESULT 4  
US-09-897-516-6783  
; Sequence 6783, Application US/09897516  
; GENERAL INFORMATION:  
; APPLICANT: Corbin, David R.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Huesling, Joseph E.  
; APPLICANT: Krasomil-Osterfeld, Karina C.  
; APPLICANT: Malvar, Thomas M.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Spiridonov, Sergei  
; TITLE OF INVENTION: Xenorhabdus sp. Genome Sequences And Uses Thereof  
; FILE REFERENCE: 38-21(51847)B  
; CURRENT APPLICATION NUMBER: US/09/897,516  
; CURRENT FILING DATE: 2001-06-29  
; PRIOR APPLICATION NUMBER: US 60/215, 161  
; PRIOR FILING DATE: 2000-06-30  
; NUMBER OF SEQ ID NOS: 8409  
; SEQ ID NO 6783  
; LENGTH: 473  
; TYPE: PRT  
; ORGANISM: Xenorhabdus sp.  
US-09-897-516-6783

Query Match 16.8%; Score 396; DB 6; Length 473;  
Best Local Similarity 31.5%; Pred. No. 1.6e-21;  
Matches 88; Conservative 70; Mismatches 109; Indels 12; Gaps 5;  
Qy 1 MAWVPSLVNGYDVAAATMAAGBMP--WMSLFGLDVAAQAGVQGVLPVLVVSIVLTIKFL 59  
Db 191 ITLVSPQLMNAIYQLQ-----QIPEWNEGFTISKVGYQAOVIFSLMAGLGHIERL 245  
Qy 60 HKRLKGTADFLLPVLTLTLTGFTTIAIGPAMRWGVDVLAHGLQGLYDFG-GPVGGLLF 118  
Db 246 KKLVPDYLVVVPVVSULLAVFLAHAIIPIGRSIGDGVAMCVKGLMTGSPAPIGAALF 305  
Qy 119 GLVYSPVITGLHQSPPIELF--NOGGSFIFATASMANIAQGAACLAFFLAKSEKIL 177  
Db 306 GFYAPLVITGVHQTALDQMIQSTGGTIPWIIALSNTAQGSAGVIGIHWASKQNER 365  
Qy 178 GLAGASGVSAVLGITEPAIFGVNLRPFETIGTAAIGGALIALENIKAVALGAAGFL 237  
Db 366 ELSVPAALISAYLGVTEPAMYGINLKYRPMFCAMIGTSLAGLLCGLNHVTSNGIVGGIP 425  
Qy 238 GVSTADPMVNFVCAVPTPIAFGAALYAGLYLVRN 276  
Db 426 GILSIKQGFQWVIYLLAMLVIIIV---PMALTVMLYRN 460

RESULT 5  
PCT-US02-03987-5658  
; Sequence 5658, Application PC/TUS0203987  
; GENERAL INFORMATION:  
; APPLICANT: Elitra Pharmaceuticals, Inc.  
; TITLE OF INVENTION: Methods for Identifying the Target of a Compound which Inhibits Q  
; FILE REFERENCE: ELITRA.028VPC  
; CURRENT APPLICATION NUMBER: PCT/US02/03987  
; CURRENT FILING DATE: 2002-02-02

; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09  
; NUMBER OF SEQ ID NOS: 15811  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 5658  
; LENGTH: 679  
; TYPE: PRT  
; ORGANISM: Staphylococcus aureus  
PCT-US02-03987-5658

Query Match 14.6%; Score 345.5; DB 1; Length 679;  
Best Local Similarity 24.2%; Pred. No. 1.4e-17;  
Matches 132; Conservative 68; Mismatches 192; Indels 153; Gaps 17;  
Qy 27 SLFGLDVAAQAGVQGVLPVLVVSIVLTIKFLHRLKRGADFL----ITPVLTLTLTG 82  
Db 123 SILGIPTLQTVFGGIIIGALAAW---CYNKFNINLSYLGFFAGKREVPIM-WATTSF 178  
Qy 83 LTFIAIGPAMRWGVDVLAHGLQ---GLYDFGPGVGLLGLVYSPVITGLHQSF--- 134  
Db 179 I--LAPPMALIW--PTIOTGLNAFSTGLLDSNTGVAVLFGFIKRLLPFGLUHFIHAPF 234  
Qy 135 -----PPIELELFNQ-----GSFIFATASMANIAQGAACLAFFFLA 171  
Db 235 WFEFGSWKNAAGEIITHGDORIFIEQIREGAHLTAGKFMQGEFPVMMFGLPAAALAIYHSA 294  
Qy 172 KSEKLKLAGAGVSA---VLGITEPAIFGVNLRWPFPIGIGTAAIGGALIALENIK 227  
Db 295 KPNKKVAVAGLSAALTSTLTGITEPLEFSFLVAPLLFFI---HAVLDGLSFLTLVLL 351  
Qy 228 AVALG---AAGFLGVVSIDA-PDMVNF-----LVCVVVTFE----- 260  
Db 352 DVHLGYTFSGGFDYVLGLVLPKNTQMWLVIPVGLVAVIYVYFVRFLVILVKLYKTPGRE 411  
Qy 261 -----AFGAALYGLY-----LVR----- 274  
Db 412 DKQSAVATASATELPYAVLEAMGKGNKIKHLDACITRLRVEVNDKSKVDVPGKLGASG 471  
Qy 275 -----RNGSIDPDATAAPVAGTAKAEAEAPFSNDST 308  
Db 472 VLEVGNNQAIQPKSDQIKHEMQQIMNGQVVENPTTMEDDKDETIVVAEDKSAITSELSH 531  
Qy 309 ITQAPLTCEAIALSVSDAMFASGLKGVATVPTKQLVSPVSGKIVVAFPGSHAFAYR 368  
Db 532 IVHADLTGVTPLSEVPDQVQVFSEKMMWGDIAIKPSQGEVRAFPNGKVQMFPTKHAIGL- 590  
Qy 369 TRAEDESNVDILMHIGFDTVNLNGTHFNPLKKGDEYKAGELLCEFDIDAIKAAGYEVTT 428  
Db 591 ---VSDSGLELLIHIGLDTVKLNGEGFTLHVEEGEVKQGDLLINFDLDYIRNHAKSDIT 647  
Qy 429 PIWVS 433  
Db 648 PIIVT 652

RESULT 6  
US-09-815-242-5658  
; Sequence 5658, Application US/09815242  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21

```

; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 5658
; LENGTH: 679
; TYPE: prt
; ORGANISM: Staphylococcus aureus
; US-09-815-242-5658

```

Query Match 14.6%; Score 345.5; DB 6; Length 679;  
Best Local Similarity 24.2%; Pred. No. 1.4e-17;  
Matches 132; Conservative 68; Mismatches 192; Indels 153; Gaps 17;

```

; APPLICANT: Xu, H. Howard
; APPLICANT: Foulkes, J. Gordon
; APPLICANT: Zamudio, Carlos
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wali, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Roemer, Terry
; APPLICANT: Jiang, Bo
; APPLICANT: Boone, Charles
; APPLICANT: Bussey, Howard
; TITLE OF INVENTION: Methods for Identifying the Target of a Compound which Inhibits the Growth of a Microorganism
; TITLE OF INVENTION: Proliferation
; FILE REFERENCE: ELITRA.028A
; CURRENT APPLICATION NUMBER: US/10/072,851
; CURRENT FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; NUMBER OF SEQ ID NOS: 15811
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5658
; LENGTH: 679
; TYPE: prt
; ORGANISM: Staphylococcus aureus
US-10-072-851-5658

```

Query Match 14.68; Score 345.5; DB 7; Length 679;  
Best Local Similarity 24.2%; Pred. NO. 1.4e-17;  
Matches 132; Conservative 68; Mismatches 192; Indels 153; Gaps 17;



QY 429 PIVS 433  
Db 650 PIVT 654

## RESULT 10

US-10-072-851-12270  
; Sequence 12270, Application US/10072851  
; GENERAL INFORMATION:  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Xu, H. Howard  
; APPLICANT: Foulkes, J. Gordon  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Roemer, Terry  
; APPLICANT: Jiang, Bo  
; APPLICANT: Boone, Charles  
; APPLICANT: Bussey, Howard  
; TITLE OF INVENTION: Methods for Identifying the Target of a Compound which Inhibits  
; FILE REFERENCE: ELITRA 028A  
; CURRENT APPLICATION NUMBER: US/10/072,851  
; CURRENT FILING DATE: 2002-02-08  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09  
; NUMBER OF SEQ ID NOS: 15811  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12270  
; LENGTH: 681  
; TYPE: PRT  
; ORGANISM: Staphylococcus aureus  
US-10-072-851-12270

Query Match 14.6%; Score 345.5; DB 7; Length 681;  
Best Local Similarity 24.2%; Pred. No. 1.4e-17;  
Matches 132; Conservative 68; Mismatches 192; Indels 153; Gaps 17;  
QY 27 SLGLDVAAGYQGTGTPVLVSVWILATIEKFLHKLKGTADFL---IPVLTLLLTGF 82  
Db 125 SILGIPTLQTVFGGIIIGALAAW---CYNKFTYNINLPVLGFFAGRFVIM-MATTSE 180  
QY 83 LTFTAGPAMRWGVDVLAHQ---GLYDFGPGVGLLFLGLVYSPVITGLHQSF--- 134  
Db 181 I--LAFPMALI--PTIQTGLNAPSTGLDSTGTVAVFLGFKRLIIPFGLHHIFHAPF 236  
QY 135 -----PPIELFLNQ-----GSIFATASMANIAQAACLAFL 171  
Db 237 WFEFGSKNAAGETIHGDQIFIEQIREGAHLTAGKFMQGEFFVMFGLPAAALAIYHSA 296  
QY 172 KSEKLKGLAGSWSA---VLGTTEPAIFGVNLRWRPFFIGTAAIGGALIALFENIK 227  
Db 297 KPNKVVAGLMGSAULTSITGITEPELSEFLFVAPLLEFI---HAVLDGLSFLTYLL 353  
QY 228 AVALG---AAGFLGVWSIDA-PDMVMP-----LVCAVVTFFI----- 260  
Db 354 DVHLGYTFSGGFDYVLLGLVLPNKTOOWLIPVGLVYAVIYVFFRLIVKLKYKTPGRE 413  
QY 261 -----AFGAATAYGLY-----LVR----- 274  
Db 414 DKSOAVTASATELPVAVLEAMGGKANIKHLDACITRLRVEVNDKSKVDVPGKLDLGASG 473  
QY 275 -----RNGSIDPDATAAPVPAGTTRKAEAPAEFSNDST 308  
Db 474 VLEGVNMQAIFGPKSQDIKHMQQIMNGOVENPTTMMDDKDETIVVVAEDKSATSELSH 533  
QY 309 IQAPLTGEAIALSSVSAMFASCKLGSGVAIVPTKGQLVSPVSGKIVVAFPPSCHAFVR 368

Db 534 IVHAPLTGEVTPLEVPDQVSEKMGDGTAKPSQSEVRAPNGKVMIPPTKHAICL- 592  
QY 369 TKAEDGSNVDILMHIGFDTVNLNTHENPLKKQDEKVGKAGELLCEFDIDAKAAGYEVTT 428  
Db 593 ---VSDSGLELLIHIGLDTVKLNGEFTLHVEEQVEKQSGDLLINFOLDYIRNHAISDIT 649  
QY 429 PIVS 433  
Db 650 PIVT 654

## RESULT 11

PCT-US02-03987-10809  
; Sequence 10809, Application PC/TUS0203987  
; GENERAL INFORMATION:  
; APPLICANT: Elittra Pharmaceuticals, Inc.  
; TITLE OF INVENTION: Methods for Identifying the Target of a Compound which Inhibits  
; FILE REFERENCE: ELITRA 028VPC  
; CURRENT APPLICATION NUMBER: PCT/US02/03987  
; CURRENT FILING DATE: 2002-02-02  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09  
; NUMBER OF SEQ ID NOS: 15811  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 10809  
; LENGTH: 484  
; TYPE: PRT  
; ORGANISM: Enterococcus faecalis  
PCT-US02-03987-10809

Query Match 14.3%; Score 338.5; DB 1; Length 484;  
Best Local Similarity 33.2%; Pred. No. 2.9e-17;  
Matches 94; Conservative 54; Mismatches 114; Indels 21; Gaps 10;  
QY 2 AMYF---PSLVNGYDVAATMAAG---EMPWMSLFGDLVQAGYQGTGTPVLVSVWILATI 55  
Db 201 ASVFGATPAL-GGVIGAVTMTGMNPDAPISNIFTGTLTSLAG-OGGIGVIFAVWLLSLL 258  
QY 56 EKFLHKLKGTADFLIIPVLTLLTGLTFTIAIGPAMRWGVDVLAHQGLQGLYDF---GG 111  
Db 259 EKOLHKVIPESIDIIVPTISLVLGLATIFLIMP---VAGAIENGLVGINVLEKGG 314  
QY 112 PVGGLLEGLVYSPVITGLHQSPPIELFLNQGG-SFIFATASMANIAQAACLAFL 170  
Db 315 MVAGFTLGLTFLPMVFMGLHQILTPHIEMINQTMGLLPLILAMAGAGQVGAALALWIR 374  
QY 171 AKSEK---LKLGLAGSWSAVIGITEPAIFGVNLRWRPFFIGTAAIGGALIALF-NIK 227  
Db 375 CKSKDKLVEMIKALPV-GILGIGEPLIYGVTLPLGRPFTTACIGGIGGAVGAFNGV 433  
QY 228 AVALGAGFLGVWSIDA-PDMVMPFLVCAVVTFFITAFGAIAIYGL 270  
Db 434 AIAIGPSGVALIPLIANQWLLAVVLGLLAAYAGGVATLFFGI 476

## RESULT 12

US-09-815-242-10809  
; Sequence 10809, Application US/09815242  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; FILE REFERENCE: ELITRA.011A



```

; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; NUMBER OF SEQ ID NOS: 15811
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10809
; LENGTH: 484
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-10-072-851-10809

Query Match      14.38;   Score 338.5;   DB 7;   Length 484;
Best Local Similarity 33.28;   Pred. No. 2.9e-17;
Matches 94;   Conservative 54;   Mismatches 114;   Indels 21;   Gaps 10;

QY      2  AMVF---PSLVNGYDVDAATMAAG---EMPWNSLFLGDVAQAQYQGTVPVLVWSWILATI 55
DB      201 ASVFGATPAL--GGVIGAVTMTGNPDAPISNIFTGGTSLAG--QGGIICVIFAWLLSLL 258

QY      56  EKFLHKKRLKGTADELITPVLPLLITGTFTTAIGPAMRWGDVLAHQGLGYDF----GG 111
DB      259  EKQLHKVPEISIDIIVTPTISLLVIGLATIFLIMP---VAGAISNGLVGIINVVLEKGG 314

QY      112 PVGGLLFGLYSPVITGLHQSFPPIELELNFQGG--SFIFATASMANTAQACLAIVFEL 170
DB      315  MVAGFTLGLTFLPAMVFGHLHOILTPIHTEMINOTGMTLLPLILAMAGAGQVCAALALWIR 374

QY      171 AKSBK--LKLIGLAGAGSVSAVIGITEPATFPGVNLRLRWPFFIGTIGTAAIGGALIALF-NIK 227
DB      375  CKSKKLVMEMIKGALPV--GIUGIGEPLIYVTLPLGRFFITACIGGGIGGAVGAGFAGNVG 433

QY      228  AVALGAAGFLGWSIDAPDMVMFLVCVAVTFFIAFGAAIAYGL 270
DB      434  AIAIGPSVALIPLIANNQWLAYVGLLAAAYAGGFVATLFFGI 476

RESULT 14
US-09-897-516-6540
; Sequence 6540, Application US/09897516
; GENERAL INFORMATION:
; APPLICANT: Corbin, David R.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Huesing, Joseph E.
; APPLICANT: Krasomil-Osterfeld, Karina C.
; APPLICANT: Malvar, Thomas M.
; APPLICANT: Slater, Steven C.
; APPLICANT: Spiridonov, Sergei
; TITLE OF INVENTION: Xenorhabdus sp. Genome Sequences And Uses Thereof
; FILE REFERENCE: 38-21(51847)B
; CURRENT APPLICATION NUMBER: US/09/897,516
; CURRENT FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/215, 161
; PRIOR FILING DATE: 2000-06-30
; NUMBER OF SEQ ID NOS: 8409
; SEQ ID NO 6540
; LENGTH: 473
; TYPE: PRT
; ORGANISM: Xenorhabdus sp.
US-09-897-516-6540

Query Match      14.18;   Score 334;   DB 6;   Length 473;
Best Local Similarity 31.88;   Pred. No. 6.1e-17;
Matches 77;   Conservative 46;   Mismatches 109;   Indels 10;   Gaps 4;

QY      2  AMVPPSLVNGYDVDAATMAAGEMPWNSLFLGDVAQAQYQGTVPVLVWSWILATIEKFLHK 61
DB      187  ALVPEMIIISAF--KASTSAGSTTEY-FIGIPLSFLNYSASVPIIFASWCKVBERKLN 243

QY      62  RLKGTADFLATPVLVTLTGTFTTAIGPAMRWGDVLAHQGLGYDFGPGVGLLFLGLV 121
DB      244  HIPATIKGEVTEPLCLILVPLVFLVIGPMATWLSOLLAGGYOAYGYVAPALAGAIMGAV 303

```



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.  
OM protein - protein search, using sw model  
Run on: March 21, 2002, 16:20:49 ; Search time 21.58 seconds  
(without alignments)  
488.023 Million cell updates/sec

Title: US-09-604-231-2  
Perfect score: 2363  
Sequence: 1 MAMVFPVLNGYDVATMAA.....IERGANLLNVAKREAVPATP 468

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	113.5	4.8	1176	1	US-07-828-788A-12
2	113.5	4.8	1176	1	US-08-356-034-2
3	113.5	4.8	1176	1	US-08-933-891-2
4	113.5	4.8	1176	5	PCT-US92-11337-12
5	113.5	4.8	1179	1	US-08-040-751-1
6	108.5	4.6	731	2	US-08-911-364-1
7	108.5	4.6	733	4	US-08-464-700-2
8	108	4.6	1011	3	US-08-836-325-2
9	108	4.6	1989	3	US-08-836-325-12
10	107	4.5	1984	3	US-08-836-325-10
11	106.5	4.5	1129	6	5164180-6
12	106.5	4.5	1179	6	5188960-2
13	106.5	4.5	1864	3	US-08-804-227C-3
14	105.5	4.5	462	3	US-08-801-344-9
15	105.5	4.5	462	4	US-09-498-599-9
16	103.5	4.4	396	2	US-08-850-880-4
17	103.5	4.4	396	2	US-08-944-916-4
18	103.5	4.4	396	2	US-08-814-877-4
19	102	4.3	3033	1	US-07-925-695-9
20	101	4.3	3033	4	US-07-925-695-8
21	100.5	4.3	522	4	US-08-894-818B-3
22	100.5	4.3	525	1	US-08-356-340-2
23	100.5	4.3	525	2	US-08-786-555-2
24	100.5	4.3	654	4	US-08-894-818B-35
25	100.5	4.3	713	3	US-09-335-409-11
26	100	4.2	528	2	US-08-403-852D-21
27	100	4.2	528	3	US-08-510-646B-22

28	100	4.2	528	4	US-09-231-818-21	Sequence 21, Appl
29	100	4.2	1969	3	US-08-836-325-16	Sequence 16, Appl
30	99.5	4.2	792	2	US-08-678-039A-40	Sequence 40, Appl
31	99	4.2	756	4	US-08-960-048-10	Sequence 10, Appl
32	99	4.2	1275	3	US-09-120-513-2	Sequence 2, Appl
33	99	4.2	1275	4	US-09-450-105-2	Sequence 2, Appl
34	98	4.1	1184	6	5254799-6	Patent No. 5254799
35	98	4.1	1188	6	5254799-7	Patent No. 5254799
36	97.5	4.1	2005	3	US-08-836-325-7	Sequence 7, Appl
37	96.5	4.1	593	6	5523211-1	Patent No. 5523211
38	96.5	4.1	648	3	US-08-810-720-13	Sequence 13, Appl
39	96.5	4.1	934	1	US-08-446-486-7	Sequence 7, Appl
40	96.5	4.1	934	1	US-08-463-308-7	Sequence 7, Appl
41	96.5	4.1	1176	1	US-08-446-486-6	Sequence 6, Appl
42	96.5	4.1	1176	1	US-08-157-363A-10	Sequence 10, Appl
43	96.5	4.1	1176	1	US-08-463-308-6	Sequence 6, Appl
44	96.5	4.1	1176	5	PCT-US93-11405A-10	Sequence 10, Appl
45	96.5	4.1	1176	5	PCT-US93-11405-10	Sequence 10, Appl

ALIGNMENTS

RESULT 1  
US-07-828-788A-12  
; Sequence 12, Application US/07828788A  
; Patent No. 5273746  
; GENERAL INFORMATION:  
; APPLICANT: PAYNE, JEWEL M.  
; APPLICANT: HICKLE, LESLIE A.  
; TITLE OF INVENTION: NOVEL BACILLUS THURINGIENSIS ISOLATES  
; TITLE OF INVENTION: ACTIVE AGAINST PHTHIRAPTERA PESTS  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: DAVID R. SALIWANCHIK  
; STREET: 2421 N.W. 41st STREET, SUITE A-1  
; CITY: GAINESVILLE  
; STATE: FL  
; COUNTRY: USA  
; ZIP: 32606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/828,788A  
; FILING DATE: 19920129  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: SALIWANCHIK, DAVID R.  
; REGISTRATION NUMBER: 31,794  
; REFERENCE/DOCKET NUMBER: MA75  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 904-375-8100  
; TELEFAX: 904-372-5800  
; INFORMATION FOR SEQ ID NO: 12:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1176 amino acids  
; TYPE: AMINO ACID  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: YES  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: BACILLUS THURINGIENSIS  
; STRAIN: AIZAWAI  
; INDIVIDUAL ISOLATE: PS811  
; IMMEDIATE SOURCE:  
; LIBRARY: LAMBDAGEM (TM) - 11 LIBRARY OF AUGUST SICK  
; CLONE: 811A2  
US-07-828-788A-12

```

Query Match          4.8%  Score 113.5;  DB 1;  Length 1176;
Best Local Similarity 22.4%;  Pred. No. 0.019;
Matches 56;  Conservative 43;  Mismatches 96;  Indels 55;  Gaps 13;

QY 249 MFLVCVVTFIAFGAAIAYGLYLVRNGSID-----PDATAAPVPAG-----TTK 294
   :|:|:  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 377 LFVLDTGTEFSASLTADLPSTIY--RQGTVDSDLVIPPDQNSVYPARAGFSHRLSHVTML 434

QY 295 AEA-----EAP-----A-EFSN---DSTIIQAPLTGEAIALSSVSDAMFASGKLGS 336
   :|:|:  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 435 SQAAGAVYTLRAPTFSWRHSAEFSLIPSSQITQIPLT-KSINLGSCTSVVKGPGFTGG 493

QY 337 GVAIVPTKGOLVSPVSGKIVVAFPSGHAFVTRKAEDGSNDVILMHIGFDVNLNGHTFN 396
   :  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 494 DILRITSPGQISTL---RVITAPLSORYRIRY--ASTTNLQFHSIDGRPINQGNFS 548

QY 397 PLKKGDEVKAGELLCFDDAIIKAAGYEVVTPVWSNYKKTGPVNTYGLGEIEAGANLL 456
   |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 549 ATMSSGGLQSG-----SFTAGF--TTPF---NFSNGSSIFTLSAHVFNESG-NEV 593

QY 457 NVAKKEAVPA 466
   :  |  |  |  |
Db 594 YIERIEFVPA 603
   :  |  |  |  |

RESULT 3
US-08-356-034-2
; Sequence 2, Application US/08356034
; Patent No. 5691308
; GENERAL INFORMATION:
; APPLICANT: Payne, Jewel M.
; APPLICANT: Sick, August J.
; TITLE OF INVENTION: No. 5691308el Bacillus thuringiensis Isolate
; TITLE OF INVENTION: Active Against Lepidopteran Pests, and Genes Encoding
; TITLE OF INVENTION: No. 5691308el Lepidopteran-Active Toxins
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: US
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/356,034
; FILING DATE:
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US/08/210,110
; FILING DATE:
; FILING DATE: 09-APR-92
; APPLICATION NUMBER: 07/865,168
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 07/451,261
; FILING DATE: 14-DEC-89
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 07/371,955
; FILING DATE: 27-JUN-89
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Saliwanchik, Roman
; REGISTRATION NUMBER: 21,023
; REFERENCE/DOCKET NUMBER: MA43.C1.D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (904)375-8100

```



GENERAL INFORMATION:  
APPLICANT: Payne, Jewel M  
APPLICANT: Sick, August J  
TITLE OF INVENTION: No. 5407825el Bacillus thuringiensis isolates  
TITLE OF INVENTION: active against Lepidopteran Pests and Genes Encoding No. 5407825el  
TITLE OF INVENTION: Lepidopteran-active Toxins  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: DAVID R. SALIWANCHIK  
STREET: 2421 N.W. 41st STREET, SUITE A-1  
CITY: GAINESVILLE  
STATE: FL  
COUNTRY: USA  
ZIP: 32606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/040,751  
FILING DATE: 19930329  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: SALIWANCHIK, DAVID R.  
REFERENCE/DOCKET NUMBER: MA39.C1.D3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 904-375-8100  
TELEFAX: 904-372-3600  
TELEX:  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1179 amino acids  
TYPE: AMINO ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: BACILLUS THURINGIENSIS  
STRAIN: AIZAWAI  
INDIVIDUAL ISOLATE: PS81I  
IMMEDIATE SOURCE:  
LIBRARY: LAMBDA GEM (TM) - 11 LIBRARY OF AUGUST SICK  
CLONE: 81R11  
PS-08-040-751-1

```

Query Match      4.8%; Score 113.5; DB 1; Length 1179;
Best Local Similarity 22.4%; Pred. No. 0.019;
Matches 56; Conservative 43; Mismatches 96; Indels 55; Gaps 13;

QY 249 MFLVCAVVTFFIAPGAIAIYGLYLVRNGSID-----PDATAAPVPAG-----TTK 294
      :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:|
Db 380 LFLVDGTEFESFASLTADLPSTIY--RQGTWVDSLVDVTPQODNSVYPARAGFSHRLSHVTML 437
      :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:|

QY 295 AEA-----EAP-----AFESN--DSTIIQAPLTGEAIALSSVSDAMFASGKLGS 336
      :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:|
Db 438 SOAGAVYTLRAPFTFSWRHRSAEFNLIPSSQITQIPLT-KSINLGSTSVVKGPGETGG 496
      :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:|

QY 337 GVAIVPTKGOLVSPVSGKIVVAPSPGSHAEAVRTKAEDGSNVDILMHICFDIVNLNGHTFN 396
      :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:|
Db 497 DIIRITSPGQISLT---RVITAPLSORYRRIHY--ASTNTQGFHTSIDGRPINQGNFS 551
      :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:|

QY 397 PLKKQGDVEKAGELLCEFDIDAKAAGAYEYVTPTWVSNRYKTKGPNVYTYGLCEIAGANLL 456
      :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:|
Db 552 ATWSSGGNLSQG-----SPTAGF--TTPF---NFSNGSSIFTLSAHFVNSG-NEV 596
      :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:|

QY 457 NVAKKEAVPA 466
      :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:|
Db 597 YIERIEFVPA 606
      :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:|

```

```

RESULT      6
US-08-911-364-1
; Sequence 1, Application US/08911364
; Patent No. 5969106
; GENERAL INFORMATION:
; APPLICANT: ROTHSTEIN, Aser
; APPLICANT: KEELY, Fred W.
; APPLICANT: ROTHSTEIN, Steven J.
; TITLE OF INVENTION: SELF-ALIGNING PEPTIDES MODELED ON HUMAN
; TYPE OF INVENTION: ELASTIN AND OTHER FIBROUS PROTEINS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY & LARDNER
; STREET: 3000 K Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/911,364
; FILING DATE: 07-AUG-1997
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/023,552
; FILING DATE: 07-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Bent, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 041082/0104
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 672-5300
; TELEFAX: (202) 672-5399
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 731 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-911-364-1

Query Match          4.6%; Score 108.5; DB 2; Length 731;
Best Local Similarity 20.6%; Pred. No 0.03;
Matches 122; Conservative 47; Mismatches 227; Indels 195; Gaps

QY   16  ATMAAGEMPWLSFLGDLVAQAQGYGTVLPLVLYSWILATIEKFLHRLKGTDADFLTIPVL 75
     | : |::| | | : | | | | | | | | | | | | | | | | | | | | | |
Db   108  AGVKPGKVGVPGLPV-----YPGGVLP-----GARFGVG----- 138

QY   76  TLLLTGFLTFIATGPAMRWGDVLNHLGQLGYDFGPGVGLLELGL-VYSPIVITGLHQSF 134
     | : | | | | | | | | | | | | | | | | | | | | | | | | | |
Db   139  --VLPGVPTGAGVKPKAPGVGAFA-GIPGVGPPGPGPGLPYGIPIKAPKLPGGYGPLY 195
     | : | | | | | | | | | | | | | | | | | | | | | | | | | |
QY   135  PPIELELFNGGGSFIFATASMANIAQG-----AACLAVFFLAKSEKLKGLAGA 182
     | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db   196  TTGKLP-YGVGPGGVAGACAGKYPTGTGVGPAAAAAATAAAKAFGAGAAGVLPVGGA 254
     | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY   183  SGVSAVLGITEPAIFGVYNLRWPFFIGITGTAIGGALIALFNFIKAVALGAAGFL----- 237
     || | | | | | | | | | | | | | | | | | | | | | | | | | |
Db   255  -GVPGVPCAI-PCIGGI-----AGVCTPAAAAAAA---AKAAKTGAAGLVPGGP 301
     | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY   238  -----GWSTSDAPDMVNF----- 251
     || | | | | | | | | | | | | | | | | | | | | | | | | | |
Db   302  GFGEVGVVPGVGVGVPAGIPVVPGAGIPGAAVPGVWSPEAAAAKAAKAAKAYCARP 361
QY   252  ---VCAVTFFIFAAGAIAYGLYLRRNG-----SIDPDATAAP--- 287

```



```

, APPLICATION NUMBER: PCT/US95/14251
, FILING DATE: 02-NOV-1995
, PRIOR APPLICATION DATA:
, APPLICATION NUMBER: 08/482,401
, FILING DATE: 07-JUN-1995
, PRIOR APPLICATION DATA:
, APPLICATION NUMBER: 08/334,029
, FILING DATE: 02-NOV-1994
, ATTORNEY/AGENT INFORMATION:
, NAME: Ludwig, Steven R.
, REGISTRATION NUMBER: 36,203
, REFERENCE/DOCKET NUMBER: 0917.0240002
, TELECOMMUNICATION INFORMATION:
, TELEPHONE: 202-371-2600
, TELEFAX: 202-371-2540
, INFORMATION FOR SEQ ID NO: 2:
, SEQUENCE CHARACTERISTICS:
, LENGTH: 1011 amino acids
, TYPE: amino acid
, TOPOLOGY: linear
, MOLECULE TYPE: protein
US-08-836-325-2

```

Query Match 4.6%; Score 108; DB 3; Length 1011;  
Best Local Similarity 20.7%; Pred. No. 0.055;  
Matches 91; Conservative 67; Mismatches 168; Indels 114; Gaps 20;

QY	28	LFGLDVAQAGYQGTVPVLVWVSWILATIEKFLHKRUKGTADFL-----ITPVLTLTLLG 81
Db	553	IFDLVNTQA-FDITIMVLICLNMTVMVEK-----EGQTEYMDYVLHWINVMVFILFTG 605
QY	82	FLTFIAIGPAMRWGVDLAHLGLQGLDFGPGVGLLFLGLVSPIVITGLHQSPFPPELEL 141
Db	606	ECVLKLIIS-----LRH-----YYP--TVG--WNILYFVVVILSVGMFLAEMIEK 646
QY	142	FNQGSFIFATASMANIAQGAACLAFFVLAKSEKLKLAGAGSVSAVLGITPEAFVGNL 201
Db	647	Y-----FVSPTLFVRILARIGRLR-----IKGAKGIRTLFLALMWSLPALFNIGL 694
QY	202	RLRWPEFI-----GIGTAA-----ICGALLALENIKAVALGAAGFLGVVS 241
Db	695	LLFLVMFIYAIFGMSFAYVKKEAGINDMFNETFGNSMICLFQIITTS-GWMDGLLAPIL 753
QY	242	IDAP-----DMVMFLVCAVVTFFIAFGAAIAYGLYLVRRNGS 278
Db	754	NSAPPCDKPKVHGGSSVEGDCGNPSVGIFYFVSYIIISFLVVNMYIA---VILENF 809
QY	279	IDPDATAAPVPAGTTKAEAEAPAESNDSTIIQAPITGEAIALSSVSDAMFASGKLGCSV 338
Db	810	VATEESTPELSEDDFMFYEVMEKEFPDAT-----QFIECKLSD-FAAALDPPLL 859
QY	339	AIVPTKGQLVS---PVSGKIIIVAPSGHAEAVRTKAEDGSNVDIL---MHIGEDTVNLN 391
Db	860	IAKPNKVQIIMDLPMVSGDRTHCLDILFATKRVLGE--GGMDSLSRQMEERFMSANPS 918
QY	392	GTHFNP-----LKKQSGDEVKA 407
Db	919	KVSYEPITITTLKRKKQEEVA 938

RESULT 9  
US-08-836-325-12  
; Sequence 12, Application US/08836325  
; Patent No. 6110672  
; GENERAL INFORMATION:  
; APPLICANT: Mandel, Gail  
; APPLICANT: Halegoua, Simon  
; APPLICANT: Borden, Laurence A.  
; TITLE OF INVENTION: Peripheral Nervous System Specific  
; TITLE OF INVENTION: Sodium Channels, DNA Encoding Therefor, Crystallization,  
; TITLE OF INVENTION: X-ray Diffraction, Computer Molecular Modeling, Rational  
; TITLE OF INVENTION: Drug Design, Drug Screening, and Methods of Making and Using

TITLE OF INVENTION: Therapeutic  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: STERNER, KESSLER, GOLDSTEIN & FOX P.L.L.C.  
STREET: 1100 New York Ave., N. W., Suite 600  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20005-3934  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/836,325  
FILING DATE: 2-MAY-1997  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/14251  
FILING DATE: 02-NOV-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/482,401  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/334,029  
FILING DATE: 02-NOV-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Ludwig, Steven R.  
REGISTRATION NUMBER: 36,203  
REFERENCE/DOCKET NUMBER: 0917.0240002  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-371-2600  
TELEFAX: 202-371-2540  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1989 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: not relevant  
MOLECULE TYPE: protein  
US-08-836-325-12

Query Match	4.6%	Score 108;	DB 3;	Length 1989;
Best Local Similarity	20.7%	Pred. No. 0.16;		
Matches 91; Conservative	67;	Mismatches 168;	Indels 114;	Gaps 20;

```

QY 28 LFLGDLVAQAGYQGTVPVLVWVSWILATIEKFLKBLKGTADFL-----ITPVLTLLLTG 81
Db 1508 IFDLVNTQA--FDITIVLICLNMVMMVEK-----EGQTEYMDVVLHWINVFIILFTG 1560
QY 82 FLTFIAGPMRWGVDLAHLGQLGYDGGPVGGLFLGLVYSPVITVGLHQSPPIELEL 141
Db 1561 ECVLKLTIS-----LRH-----YYF--TVG---WNLYFVVVILSVGMFLAEMIEK 1601
QY 142 FNQGGSFIFATASMANIAGAACLAFFVLAKSEKLKGLAGASVAVLGITEPAIFGVNL 201
Db 1602 Y-----FVSPFLFVIRIARIGRLRL-----IKGAKGIRTLFLFALMSLPALFNIGL 1649
QY 202 RLRWPEFFI-----GIGTAA-----ICGALIALFNKAVAGAACGFLGWVS 241
Db 1650 LFLVWFIVAIFGMSFAYVKKEAGINDMFNETPGNSMICLFOITISA--GWDGLLAPIL 1708
QY 242 IDAP-----DMVFLCAVVTFFIARFGAATAVGLYLVRNGS 278
Db 1709 NSAPPCDPKKVHPGSSVGGDCGNPSVGIFVSVYIIISFLVWVMVIA-----VILENFS 1764
QY 279 IDPDATAAPVPAGTTKAEAEAPAESNDSTIIQAPLTCEATLSVSDAMPASCKLGSV 338
Db 1765 VAITESTEPLSEDDPFEMFYVEWKEKDDPAT-----QFTEFKCLSD--FAALADPPLL 1814
QY 339 AIVPTKGLQVS-----PVSCKIVVAPSPGHAFVTRKAEODGNSVDIL---MHIGFDTVNLN 391

```



Db 1815 IAKPNKVLQIAMDLPVSGDRIHCLDILFAFTRKVLGE--GGEMDSLRSQMEERFMSANPS 1873  
QY 392 GTHFNP-----LKKQGEVKA 407  
Db 1874 KVSYPEITTTTLKRKQEVSA 1893  
RESULT 10  
US-08-836-325-10  
; Sequence 10, Application US/08836325  
; Patent No. 6110672  
; GENERAL INFORMATION:  
; APPLICANT: Mandel, Gail  
; APPLICANT: Haleboua, Simon  
; APPLICANT: Borden, Laurence A.  
; TITLE OF INVENTION: Peripheral Nervous System Specific  
; TITLE OF INVENTION: Sodium Channels, DNA Encoding Therefor, Crystallization,  
; TITLE OF INVENTION: X-ray Diffraction, Computer Molecular Modeling, Rational  
; TITLE OF INVENTION: Drug Design, Drug Screening, and Methods of Making and Using  
; TITLE OF INVENTION: Thereof  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C  
; STREET: 1100 New York Ave., N. W., Suite 600  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20005-3934  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/836,325  
; FILING DATE: 2-MAY-1997  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US95/14251  
; FILING DATE: 02-NOV-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/482,401  
; FILING DATE: 07-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/334,029  
; FILING DATE: 02-NOV-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ludwig, Steven R.  
; REGISTRATION NUMBER: 36,203  
; REFERENCE/DOCKET NUMBER: 0917.024.0002  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-371-2600  
; TELEFAX: 202-371-2540  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1984 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-836-325-10

Query Match 4.5%; Score 107; DB 3; Length 1984;  
Best Local Similarity 20.7%; Pred. No. 0.2;  
Matches 91; Conservative 66; Mismatches 169; Indels 114; Gaps 20;  
QY 28 LFGLDVAQAGYQGTVPVLVWVSHILATIEKFLKRLKGTADFL-----ITPVLTLTLTG 81  
Db 1506 IFDLVTNQA-FDITIMVLCLNMTVMVEK-----EGQTEYMDYVLHWINMVFILFTG 1558  
QY 82 FLTFIAIGPAMRWGVDLAHLGLOGLYDFGPGVGLFLGLVYSPIVITGLHQSFPPILEL 141

Db 1559 ECVLKLIS-----LRH-----YVF--TVGWNIEFVVVILSVGMFLA-----EM 1596  
QY 142 FNOGGSEIFATASWANTAOGAACLAFFLAKSEKLGAGASGVSAVLGITPAIFGVNL 201  
Db 1597 IEK-----YFVSPTLFRVIRLARIGRIURL-----IKGAKGIRTLFLFALMMSLPALFNIGL 1647  
QY 202 RLWPFPI---GIGTAA-----IGGALIALFNKAKAVALCAAGFLGVWS 241  
Db 1648 LLFLVFIYAIFGMSNFAYVYKKEAGINDMFENFETFGNSMCLFOITTSAGWDGULLAPIL 1706  
QY 242 IDAP-----DMVNFVCAVVVTFEFAFGAAIAYGLYLVRNRS 278  
Db 1707 NSAPPDCDPKXVHPGSSVEGDCGNPSVGFVFSYIIISFLVVNMYIA-----VILENFS 1762  
QY 279 IDPDATAAPVPAGTTKAEAPAEFSNDSTIIQAPLTGEAIALSSVDAMFASCKLGSV 338  
Db 1763 VATEESTPELSEDDFEMFYEWEKFPDAT-----QFIEFCKLSD--FAAALDPPLL 1812  
QY 339 AIVPTKGQLVS---PVSGKIVVAFPSGHAFVARTKAEDGNSVDIL---MHIGFDTVNLN 391  
Db 1813 IAKPNKVLQIAMDLPVSGDRIHCLDILFAFTRKVLGE--GGEMDSLRSQMEERFMSANPS 1871  
QY 392 GTHFNP-----LKKQGEVKA 407  
Db 1872 KVSYPEITTTTLKRKQEVSA 1891  
RESULT 11  
5164180-6  
; Patent No. 5164180  
; APPLICANT: Payne, Jewel; Sick, August J.  
; TITLE OF INVENTION: BACILLUS THURINGIENSIS ISOLATES ACTIVE  
; AGAINST LEPIDOPTERAN PESTS  
; NUMBER OF SEQUENCES: 6  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/451,389  
; FILING DATE: 14-DEC-1989  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 353,860  
; FILING DATE: 18-MAY-1989  
; SEQ ID NO: 6;  
; LENGTH: 1129  
5164180-6  
Query Match 4.5%; Score 106.5; DB 6; Length 1129;  
Best Local Similarity 22.4%; Pred. No. 0.093;  
Matches 56; Conservative 42; Mismatches 97; Indels 55; Gaps 13;  
QY 249 MELVCAVVTFFIAFGAAIAYGLYLVRNRSID-----PDATAAPVPAG-----TTK 294  
Db 380 LFLVDGTEFFASLTADLPSTIY--RORGIVDSLDVIPPQDNSVPPARAGFSHRLSHVTML 437  
QY 295 AEA-----EAP-----AEFSN---DSTIIQAPLTGEAIALSSVDAMFASCKLGS 336  
Db 438 SQAAGAVYTLRAPTFSWRHSAEFSNLIPSSQITQIPLT-KSINLGSCTSVVKGPGGTGG 496  
QY 337 GVAIVPTKGOLVSPVSKIVVAFPSGHAFVARTKAEDGNSVDILMHIGFDTVNLNTHFN 396  
Db 497 DILRRTPSGQISTL---RVITITAPLSQRVRYRY---ASTTNLQFHTSIDGRPNQGNFS 551  
QY 397 PLKKQGEVKAAGELLCEFDIDAIAKAGYEVTTPIVVSNNYKKTGPVNTYGLGEIAGANLL 456  
Db 552 ATWSSGGNLSQSG-----SFTAGF--TTPF---NFSNGSSIFTLSAHVFNSG-NEV 596  
QY 457 NVAKKEAVPA 466  
Db 597 YIDRIEFVPA 606  
RESULT 12  
5188960-2  
; Patent No. 5188960

; APPLICANT: PAYNE, JEWEL; SICK, AUGUST J.  
; TITLE OF INVENTION: BACILLUS THURINGIENSIS ISOLATE ACTIVE  
; AGAINST LEPIDOPTERAN PESTS, AND GENES ENCODING NOVEL  
; LEPIDOPTERAN-ACTIVE TOXINS  
; NUMBER OF SEQUENCES: 8  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/451,261  
; FILING DATE: 14-DEC-1989  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 371,955  
; FILING DATE: 27-JUN-1989  
; SEQ ID NO:2:  
; LENGTH: 1179  
5188960-2

Query Match 4.5%; Score 106.5; DB 6; Length 1179;  
Best Local Similarity 22.4%; Pred. No. 0.1;  
Matches 56; Conservative 42; Mismatches 97; Indels 55; Gaps 13;  
QY 249 MFLCAVVTETIAGAAATAYGLYLVRNGSID-----PDATAAPVAG-----TTK 294  
Db 380 LFVLDGTEFFASLTADLPSTIY--RORGTVDSLDVIPPQNSVPARAGFSHRLSHVTML 437  
QY 295 AEA-----EAP-----AEFSN---DSTIIQAPLTGEAIALSSVSDAMFASGKLGS 336  
Db 438 SOAGAVYTLRAPTFMSRHSRAEFSNLIPSSQITQIPLT-KSINLGSCTSVKPGFTGG 496  
QY 337 GVALVPTKQLVSPVSKIVVAFPSGHAFVTRKAEDGNSVDILMHIGFDVFNLNHFN 396  
Db 497 DILRTSPGQISTL---RVITITAPLSQRYRIRY--ASTTNLQFHTSIDGRPINQGNFS 551  
QY 397 PLKQGEVAGELLCEFDIDAIAKAGVEVTPIVVSNYKKTGPVNYGLGEIEAGANLL 456  
Db 552 ATMSSGNLQSG-----SFRTAGF--TTFP---NFSNGSIFTLSAHFVNSG-NEV 596  
QY 457 NVAKKEAVPA 466  
Db 597 YIDRIEFVPA 606

RESULT 13  
US-08-804-227C-3  
; Sequence 3, Application US/08804227C  
; Patent No. 5876991  
; GENERAL INFORMATION:  
; APPLICANT: Dehoff, Bradley S.  
; APPLICANT: Kunstoss, Stuart A.  
; APPLICANT: Rostock, Paul R., Jr.  
; APPLICANT: Sutton, Kimberly L.  
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: THOMAS G. PLANT 1501  
; STREET: LILLY CORPORATE CENTER  
; CITY: INDIANAPOLIS  
; STATE: IN  
; COUNTRY: USA  
; ZIP: 46285  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: ASCI(DOS) Text only  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/804,227C  
; FILING DATE: February 21, 1997  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Plant, Thomas G.  
; REGISTRATION NUMBER: 35,784  
; REFERENCE/DOCKET NUMBER: X-8231  
; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 317-276-2459  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1864 amino acids  
; TYPE: amino acid  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
US-08-804-227C-3

Query Match 4.5%; Score 106.5; DB 2; Length 1864;  
Best Local Similarity 21.7%; Pred. No. 0.2;  
Matches 94; Conservative 45; Mismatches 159; Indels 135; Gaps 20;  
QY 7 SLVNGYDVATMAAGEMPW-----SLGLDVAAGYOGTVLPVLVSVNIIATIEKFLHK 61  
Db 346 ALLATYGTART---AERPLWLGLSKSNIGHTQAAAGVAGVIMVLMAR-----HG 392  
QY 62 RL-----KGTADFLLITPVLTLTLLTGLTFTTFAIGPAMRWGVDVLAHGLQG 105  
Db 393 RLPRTLHVDRTTRVDWEKGGVRLTETV-----PWPGEAGEPRRAG 434  
QY 106 LYDFGGPVGGLFLGLVSPVITGLHQSFPIELFNOGGSFIFATASMANIAQGAACL 165  
Db 435 VSSEGA-----SGTNAHVVLVSPAGEPPAAGRPEDTGAW---TVS---GRGPAAL 480  
QY 166 AVFLAKSEKLGKLAGASGVSAVLGITEPAIFGVNLRRLRWPFFIGTGTAAIGGALI---A 222  
Db 481 RAQAARLYDALTGTTGTGGAGOG-----AGPGTAEGVAGALAHART 522  
QY 223 LFNIKAVALGA-----AGFLGVVVSIDAPDMVMFLVCAVW---TFFTAFG---AAIA 267  
Db 523 AFRRAVAVLGNRAELLAGLRELAEEHPGPRVVTGTATPATERRTAFLEFSQGSORAGS 582  
QY 268 YGLYLVRN-----GSDPD-----ATAAPVPAGTTKAEAEAPAEFSN 305  
Db 583 RGLY--RRHPVFAALDEVCAALEPHLRPLRDLMFABPGSPAEPLDRTFTQPALEAL 640  
QY 306 DSTIIQAP---LTGEAIALSSVSDAMFASGKLGSVAIVTKGOLVSPVSGKIVVAPPS 361  
Db 641 QTALFRLAEHHGLRAELCGHSGVEIAAAH---AAGVLTLPLDAARLVA-ARGRLQMALPA 696  
QY 362 GHAF-AVTRKAEAD 373  
Db 697 GGAAALRATAEE 709

RESULT 14  
US-08-801-344-9  
; Sequence 9, Application US/08801344  
; Patent No. 6087140  
; GENERAL INFORMATION:  
; APPLICANT: Cameron, Douglas C.  
; APPLICANT: Shaw, Anita J.  
; APPLICANT: Altaras, Nedim E.  
; TITLE OF INVENTION: MICROBIAL PRODUCTION OF 1,2-PROPANEDIOL  
; TITLE OF INVENTION: FROM SUGAR  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dewitt Ross & Stevens S.C.  
; STREET: 8000 Excelsior Drive, Suite 401  
; CITY: Madison  
; STATE: WI  
; COUNTRY: U.S.A.  
; ZIP: 53717-1914  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patencin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/801,344  
; FILING DATE:

ATTORNEY/AGENT INFORMATION:  
 NAME: Sara, Charles S.  
 REFERENCE/DOCKET NUMBER: 09820.037  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 608-831-2100  
 TELEFAX: 608-831-2106  
 INFORMATION FOR SEQ ID NO: 9:

```

; LENGTH: 402 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Pyridine nucleotide transhydrogenase, subunit
; ORGANISM: B
; US-09-498-599-9

```

Query Match	4.5%;	Score 105.5;	DB 4;	Length 462;
Best Local Similarity	21.1%;	Pred. No. 0.029;		
Matches	91;	Conservative 66;	Mismatches 123;	Indels 151; Gaps
Qy	27	SLFGDVAQAQY----	---QGTVLPLV----	---VSWILATIEKFL 59
		: : : : : :	: : : : : :	: : : : : :
Db	48	TIFGPDGTGNVILLAMVIGGAIGIRLAKKVVEMTEMLVAILHSFVGLAALVGFNSYL 107		
Qy	60	HKRLKGTADFLITPVLTLLITGTLFTTAIGPARWGDVLAHGLOGLYDGGPVGGLFG 119		
		: : : :   :	: : : :   :	: : : :   :
Db	108	HHD--AGMAPILVNIIHLEVEFLG----	---IFIG--AVTFTGSVAF-----	---GKLGC 148
Qy	120	LVS--PIVITGLHGSFPPIELELFGNOGGSFTATASMANIAQGAACIAVFFLAKSEKLG 178		
		: : : :   :	: : : : : :	: : : : : :
Db	149	KISKPLMLPNRH-----	-----MNLAAVLVSFLLLIIVFVRTDSVGLQ--	187
Qy	179	LAGASGVSVLGTTEPA--IFGVNLRLRWPFFIGTGAICGA-----	---LIALEN--IKAAVLG 232	232
		: : : :   :	: : : :   :	: : : :   :
Db	188	-----VLIALLIMTAIALVFGWHL-----	-----VASIGADMPVVVSLNLSYSGWAAA 230	230
Qy	233	AAGFL-----	---GVYSIDAPDMVMEVLCAVVTFFTAFGAAITAGLYLVRRNGSIDPDA 283	283
		:	: : : : : :	: : : : : :
Db	231	ACGFMISNDIILVTGALVGSAGATISVIMCKANN--PSFTSVTAGGP-----	---GTDGSS 281	281

[illegible]

Search completed: March 21, 2002, 16:23:16  
Job time: 147 sec.

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 21, 2002, 16:19:04 ; Search time 42.58 Seconds  
(Without alignments)  
814.145 Million cell updates/sec

Title: US-09-604-231-2  
Perfect score: 2363  
Sequence: 1 MAMVPSLVNGYDVATMAA.....IEAGNLLNVAKKAVPATP 468

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_1101.\*  
1: /SID88/gcgdata/geneseq/geneseq/AA1980.DAT.\*  
2: /SID88/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
3: /SID88/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
4: /SID88/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
5: /SID88/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
6: /SID88/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
7: /SID88/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
8: /SID88/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
9: /SID88/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
10: /SID88/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
11: /SID88/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
12: /SID88/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
13: /SID88/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
14: /SID88/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
15: /SID88/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
16: /SID88/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
17: /SID88/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
18: /SID88/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
19: /SID88/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
20: /SID88/gcgdata/geneseq/geneseq/AA1999.DAT.\*  
21: /SID88/gcgdata/geneseq/geneseq/AA2000.DAT.\*  
22: /SID88/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	2363	100.0	468	22	AAB66707 C-glutamicum phosph
2	2363	100.0	661	22	AAG92650 C glutamicum prote
3	2332	98.7	661	22	AAB69080 Brevibacterium lac
4	1821	77.1	362	22	AAB66708 C-glutamicum phosph
5	468.5	19.8	683	22	AAG93207 C glutamicum prote
6	468.5	19.8	683	22	AAB66721 C-glutamicum phosph
7	298.5	12.6	648	22	AAG98282 Escherichia coli p
8	234	9.9	135	22	AAB66723 C-glutamicum phosph
9	233.5	9.8	167	20	AA190012 B. burgdorferi ant
10	232.5	9.8	583	21	AA153620 Amino acid sequenc
11	166.5	7.0	135	20	AA120013 B. burgdorferi ant

12	166.5	7.0	413	22	AAB66722 C-glutamicum phosph
13	161.5	6.8	126	18	AAW28051 Amino acid sequenc
14	156	6.6	449	21	AA181556 Streptococcus pneu
15	154	6.5	108	21	AAB40791 Human ORFX ORF555
16	151.5	6.4	658	21	AAB15995 E. coli proliferat
17	137	5.8	572	21	AA181595 Streptococcus pneu
18	134	5.7	423	22	AAG83033 S. epidermidis ope
19	133	5.6	48	18	AAW28158 Amino acid sequenc
20	128.5	5.4	806	20	AA121797 Alteromonas lipase
21	128.5	5.4	809	20	AA121796 Alteromonas lipase
22	128	5.4	629	22	AAG92331 C glutamicum prote
23	128	5.4	629	22	AAB76821 Corynebacterium gl
24	118.5	5.0	805	22	AAG90449 C glutamicum prote
25	118	5.0	461	20	AA138792 Neisseria meningit
26	117	5.0	461	20	AA138791 Neisseria meningit
27	116	4.9	397	22	AAW96600 Corynebacterium gl
28	116	4.9	502	20	AAW99600 B-stationis multi-
29	116	4.9	506	21	AA181620 Streptococcus pneu
30	115	4.9	521	21	AA13578 Streptomycetes globi
31	115	4.9	521	21	AA13605 Streptomycetes globi
32	114	4.8	461	20	AA138793 Neisseria gonorrhoe
33	114	4.8	461	20	AA138794 N. gonorrhoeae ant
34	113.5	4.8	1176	14	AA139755 Delta endotoxin.
35	112	4.7	1951	22	AA199678 Human adult form o
36	112	4.7	1951	22	AA199679 Human neonatal for
37	111	4.7	551	19	AAW98519 H. pylori GPO 132
38	110.5	4.7	528	22	AAW38697 Human polypeptide
39	110.5	4.7	581	22	AAW40483 Human polypeptide
40	109	4.6	3033	14	AA135339 NAMH virus strain
41	108.5	4.6	730	19	AAW46315 Human elastin cont
42	108.5	4.6	731	21	AA169068 Amino acid sequenc
43	108.5	4.6	731	22	AA166657 Human elastin prot
44	108.5	4.6	733	15	AA156653 Synthetic human tr
45	108.5	4.6	733	20	AA151301 Amino acid sequenc

ALIGNMENTS

RESULT 1  
AAB66707  
ID AAB66707 standard; protein; 468 AA.  
XX  
AC AAB66707;  
XX  
DT 09-APR-2001 (first entry)  
XX  
DE C-glutamicum phosphoenolpyruvate protein #1.  
XX  
KW Phosphoenolpyruvate; sugar phosphotransferase system; PTS.  
XX  
OS Corynebacterium glutamicum.  
XX  
PN WO200102583-A2.  
XX  
PD 11-JAN-2001.  
XX  
PF 27-JUN-2000; 2000WO-IB00973.  
XX  
PR 01-JUL-1999; 99US-0142691.  
PR 23-AUG-1999; 99US-0150310.  
PR 03-SEP-1999; 99DE-1042095.  
PR 03-SEP-1999; 99DE-1042097.  
XX  
(BADI ) BASF AG.  
XX  
PI Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;  
XX  
DR WPI; 2001-080989/09.  
XX  
PT Corynebacterium glutamicum nucleic acids encoding phosphoenolpyruvate;  
PT sugar phosphotransferase system proteins or their portions, useful for  
PT typing or identifying C. glutamicum or related bacteria, and as markers

PT for transformation -

XX Claim 4; Page 101-102; 144pp; English.  
XX  
CC The present invention relates to Corynebacterium glutamicum  
CC phosphoenolpyruvate: sugar phosphotransferase system (PTS) proteins.  
CC The PTS nucleic acids and proteins are useful in the  
CC identification of microorganisms which can be used to produce fine  
CC chemicals, for modulating fine chemical production in C. glutamicum or  
CC related bacteria, the typing or identification of C. glutamicum or  
CC related bacteria, as reference points for mapping C. glutamicum genome,  
CC and as markers for transformation.  
XX  
SQ Sequence 468 AA;

Query Match 100.0%; Score 2363; DB 22; Length 468;  
Best Local Similarity 100.0%; Pred. No. 7.7e-228;  
Matches 468; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAMVPSLVNGYDVAAATMAAGEMPMSLFGDLVAQAGYQGTVPVLVSVILATIEKFLH 60  
DB 1 mamvfslvngydvaaatmaagempmslfgdlvaqagyggtvlpvlsvvillatiekflh 60  
QY 61 KRKLGTADEFLITPVLTLLLTGFTLFTTAIGPAMRWGVDVLAHQGLQGLYDFGPGVGGLLFGL 120  
DB 61 krlkgtadflitpvtlllltfgftlfttaigpamrwvgdvlahgllqglydfgpgvggllfgl 120  
QY 121 VYSPVITGLHQSFPPPIELFNOGGSFFIPATASMANIAGAACLVFFIAKSEKLGKLA 180  
DB 121 vvspvittglhqsfpplielefnoggsffipatasmaniaagaaclavfflaksekikgla 180  
QY 181 GASGVSALVIGITEPAIFGVNLRWRPFFIGTAAIGGALIALFNKAVAGFVGVV 240  
DB 181 gasgvsavlitgitepaifgvnlrwrpffigttaaiggallialfnkavagfvgvv 240  
QY 241 SIDAPDMVFLCAVVTFFIAFGAAIAYGLYLVRRNGSIDPDATAAPVPAGTTKAEAP 300  
DB 241 sidapdmvflcavvtffiafgaaiaayglylvrrngsidpdataapvpagttkaeap 300  
QY 301 AEFNSNSTIIQAPLTGEATIALSSVSDAMFASGKLGSGVAIVPTKGQLVSPVSKIVVAFP 360  
DB 301 aefnsndstiiqapltgeatials svsdamfasklgsgvaivptkqglvsvpskivvafp 360  
QY 361 SGHAFVTRKAEQGSNVDILMHIGFDTVNLNGTHFNPLKQGDVKAAGELLCFEDDAIK 420  
DB 361 sghafavtrkaedgsnvdilmhigfdtnlngthfnplkqgdvkaagellcfeddaik 420  
QY 421 AAGYEVTTPVIVSNYKKTGPVNTYGLGEIDAGANLLNVAKKEAVPATP 468  
DB 421 aagyevttpvivsnkktgpvntyglgeidaganllnvakkeavpatp 468

RESULT 2  
AAG92650  
ID AAG92650 standard; Protein; 661 AA.

XX  
AC AAG92650;  
XX  
XX 26-SEP-2001 (first entry)  
DT

DE C glutamicum protein fragment SEQ ID NO: 6404.  
XX  
KW Corynebacterium; amino acid synthesis; vitamin; saccharide;  
KW organic acid synthesis.

XX  
OS Corynebacterium glutamicum.  
XX  
PN EP1108790-A2.

XX  
PD 20-JUN-2001.  
XX  
PF 18-DEC-2000; 2000EP-0127688.

XX

PR 16-DEC-1999; 99JP-0377484.  
PR 07-APR-2000; 2000JP-0159162.  
PR 03-AUG-2000; 2000JP-0280988.

XX  
PA (KYOW ) KYOWA HAKKO KOGYO KK.

XX  
PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;  
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;

XX  
DR WPI; 2001-376931/40.

XX  
DR N-PSDB; AAH67869.

XX  
PT Novel polynucleotides derived from Coryneform bacteria, for identifying  
PT mutation point of a gene, measuring expression of a gene, analysing  
PT expression profile or pattern of a gene and identifying homologous gene  
PT

PS Claim 17; SEQ ID NO: 6404; 246pp + Sequence Listing; English.

XX The present invention provides a number of nucleotide and protein  
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These  
CC are useful for identifying the mutation point of a gene derived from a  
CC mutant of coryneform bacterium, measuring expression amount and  
CC analysing the expression profile or expression pattern of a gene derived  
CC from Coryneform bacterium, and identifying a homologue of a gene derived  
CC from coryneform bacterium. Coryneform bacteria are useful for producing  
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,  
CC particularly L-lysine. The present sequence is a protein described  
CC in the exemplification of the invention.

CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from the  
CC European Patent Office.

XX Sequence 661 AA;

Query Match 100.0%; Score 2363; DB 22; Length 661;  
Best Local Similarity 100.0%; Pred. No. 1.3e-227;  
Matches 468; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAMVPSLVNGYDVAAATMAAGEMPMSLFGDLVAQAGYQGTVPVLVSVILATIEKFLH 60  
DB 194 mamvfslvngydvaaatmaagempmslfgdlvaqagyggtvlpvlsvvillatiekflh 253

QY 61 KRKLGTADEFLITPVLTLLLTGFTLFTTAIGPAMRWGVDVLAHQGLQGLYDFGPGVGGLLFGL 120  
DB 254 krlkgtadflitpvtlllltfgftlfttaigpamrwvgdvlahgllqglydfgpgvggllfgl 313

QY 121 VYSPVITGLHQSFPPPIELFNOGGSFFIPATASMANIAGAACLVFFIAKSEKLGKLA 180  
DB 314 vvspvittglhqsfpplielefnoggsffipatasmaniaagaaclavfflaksekikgla 373

QY 181 GASGVSALVIGITEPAIFGVNLRWRPFFIGTAAIGGALIALFNKAVAGFVGVV 240  
DB 374 gasgvsavlitgitepaifgvnlrwrpffigttaaiggallialfnkavagfvgvv 433

QY 241 SIDAPDMVFLCAVVTFFIAFGAAIAYGLYLVRRNGSIDPDATAAPVPAGTTKAEAP 300  
DB 434 sidapdmvflcavvtffiafgaaiaayglylvrrngsidpdataapvpagttkaeap 493

QY 301 AEFNSNSTIIQAPLTGEATIALSSVSDAMFASGKLGSGVAIVPTKGQLVSPVSKIVVAFP 360  
DB 494 aefnsndstiiqapltgeatials svsdamfasklgsgvaivptkqglvsvpskivvafp 553

QY 361 SGHAFVTRKAEQGSNVDILMHIGFDTVNLNGTHFNPLKQGDVKAAGELLCFEDDAIK 420  
DB 554 sghafavtrkaedgsnvdilmhigfdtnlngthfnplkqgdvkaagellcfeddaik 613

QY 421 AAGYEVTTPVIVSNYKKTGPVNTYGLGEIDAGANLLNVAKKEAVPATP 468

DB 614 aagyevttpvivsnkktgpvntyglgeidaganllnvakkeavpatp 661



Db 1 ydfgppvgllflgvspivltghqsfppielelfnfggggsffifatasmanlagaacla 60  
 QY 167 VFFLAKEKLGKAGASVAVLGTEPAIEFVNLRLRWPFEGTGTAAIGGALIALFNI 226  
 Db 61 vfflaksekklagagvsavlgitepaifgvnlrlrwpffigtaagallafni 120  
 QY 227 KAVALGAAGFLGVYSIDAPDMVFLVCAVTFVFFIAFGAAIAYGLYLVRNGSIDPDATAA 286  
 Db 121 kavalgaagflgvysidapdmvmfvcavtffiafgaaiaayglylvrrngsidpataa 180  
 QY 297 PVPAGTTKAEAEAPAEFSNDSTIIOTAPLTGEAIALSSVSDAMFASGKLGSGVAIVPTKGQ 346  
 Db 181 pvpagttkaeaeapaeFSNDSTIIOTAPLTGEAIALSSVSDAMFASGKLGSGVAIVPTKGQ 240  
 QY 347 LVSPVSGKIVVAFPSGHAFVTRKAEDGSNVDILMHIGFDVNLNGHFNPLKKQGEVK 406  
 Db 241 lvspvsgkivvafpsghafvtrkaedgsnvdilmhigfdvnlngthfnplkkqgevk 300  
 QY 407 AGELLCFDFIDAIKAAGYEVVTPPIVVSNNYKKTGPVNTYGLGEIEAGANLLNVAKKEAVPA 466  
 Db 301 agellicfdidaikaagyevtppivvsnnnykktgpvntyglgeieaganllnvakkeavpa 360  
 QY 467 TP 468  
 Db 361 tp 362

RESULT 5  
 AAG93207  
 ID AAG93207 standard; Protein; 683 AA.  
 XX AC AAG93207;  
 XX DT 26-SEP-2001 (first entry)  
 XX DE C glutamicum protein fragment SEQ ID NO: 6961.  
 XX KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;  
 XX KW organic acid synthesis.  
 XX OS Corynebacterium glutamicum.  
 XX PN EP1108790-A2.  
 XX PD 20-JUN-2001.  
 XX PF 18-DEC-2000; 2000EP-0127688.  
 XX PR 16-DEC-1999; 99JP-0377484.  
 XX PR 07-APR-2000; 2000JP-0159162.  
 XX PR 03-AUG-2000; 2000JP-0280988.  
 XX PA (KYOW ) KYOWA HAKKO KOGYO KK.  
 XX PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;  
 XX PI Tateishi N, Senoh A, Ikeda M, Ozaki A;  
 XX DR WPI: 2001-376931/40.  
 XX DR N-PSDB; AAH68426.  
 XX PT Novel polynucleotides derived from Coryneform bacteria, for identifying  
 XX PT mutation point of a gene, measuring expression of a gene, analysing  
 XX PT expression profile or pattern of a gene and identifying homologous gene  
 XX PS Claim 29; SEQ ID NO: 6961; 246pp + Sequence Listing; English.  
 XX CC The present invention provides a number of nucleotide and protein  
 XX CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These  
 XX CC are useful for identifying the mutation point of a gene derived from a  
 XX CC mutant of coryneform bacterium, measuring expression amount and  
 XX CC analysing the expression profile or expression pattern of a gene derived  
 XX CC from Coryneform bacterium, and identifying a homologue of a gene derived

CC from coryneform bacterium. Coryneform bacteria are useful for producing  
 CC amino acids, nucleic acids, vitamins, saccharides and organic acids,  
 CC particularly L-lysine. The present sequence is a protein described  
 CC in the exemplification of the invention.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from the  
 CC European Patent Office.

XX Sequence 683 AA;

Query Match 19.8%; Score 468.5; DB 22; Length 683;  
 Best Local Similarity 29.8%; Pred. No. 6.3e-38;  
 Matches 141; Conservative 79; Mismatches 206; Indels 47; Gaps 9;

QY 27 SLFGLDVAQAGYQGTVPVLVSVILATIERFLHKLKGTADFLITPVLTLTLLGFLTFPI 86  
 Db 218 tvfglpmvlnndysgvpfppllaaiglywvekglkklpeavgmfvfssllimipataf 277  
 QY 87 AIGPAMRWGVDVLAHGLQGLYDFGPGVGLLFGVYSPIVITGLHOSFPPIELELFNQ-G 145  
 Db 278 llgpfvgvngisnllleainmfslisvlpilypflvplghlwnpnaimiqnintlg 337  
 QY 146 GSFIFATASMANIAQGAACLAFFFLAKSEKLGLAGAS--GVSAVL--GITEPAIFGVNL 201  
 Db 338 ydfigppmgawnfacgltvgtvfillsikernkamrqvslgmlagllggisepslygvll 397  
 QY 202 RLRPFFIGTGTAAIGGALIALFNKKAVALGAAGFLGVVSDADPDWMLVCVVTFPIA 261  
 Db 398 rfkktvfrllpgclagvgimgidikayafvtslltipamd--pwlgytigiavafivs 455  
 QY 262 FGAAIAYGLYLVRNGSID-----PDAT-AAPV----- 288  
 Db 456 mflviald---yrsneerdearakaadkqaedlkaeanatpaapvaagagagaga 512  
 QY 289 PAGTTKAEAEAPAEFSNDSTIIQAPLTGEAIALSSVSDAMFASGKLGSGVAIVPTKGOLV 348  
 Db 513 aagaataavaakpklaagevvdvpslegkaiplsevpdpifaegklpggiaiqptgntvv 572  
 QY 349 SPVSGKIVVAFPSGHAFVTRKAEDGSNVDILMHIGFDVNLNGHFNPLKKQGEVKAG 408  
 Db 573 apadatvllvqksghavairl-----dagveillvhvgldvtqlggegftvhvrrqqvkag 628  
 QY 409 ELLCEFDIDAICAAGYEVVTPPIVVSNNYKKTGPVNTYGLGEIEAGANLLNVAKK 461  
 Db 629 dplitfdadfirskdlpitpvvvsnaakfgeiegipadqansstvtikvngk 681

RESULT 6

AAB66721  
 ID AAB66721 standard; protein; 683 AA.

XX AC AAB66721;

XX DT 09-APR-2001 (first entry)

XX DE C.glutamicum phosphoenolpyruvate protein #5.

XX KW Phosphoenolpyruvate; sugar phosphotransferase system; PTS.

XX OS Corynebacterium glutamicum.

XX PN WO200102583-A2.

XX PD 11-JAN-2001.

XX PF 27-JUN-2000; 2000WO-IB00973.

XX PR 01-JUL-1999; 99US-0142691.

XX PR 23-AUG-1999; 99US-0150310.

XX PR 03-SEP-1999; 99DE-1042095.

XX PR 03-SEP-1999; 99DE-1042097.





Qy 221 -----IALFNKA-----VALGAA- 234  
Dy 345 viffayfvfslvirmfnlktptgredkedelivteeansteegltqlatnyiaavgtd 404  
Qy 235 -----GFLGVSTIDAPDMWFLVCAVVTFFIAFGA 264  
Dy 405 nlkaidacitrlritvadsarvndtmckrlgagvkvlnkqtiqivgakaesigdamkk 464  
Qy 265 AIAYGLVLRNRGSDPDATAAPVPAGTTKAEAEAPAEFSNDSTIIQAPLTGEATLSSV 324  
Dy 465 vvar9-pvaasaeatp-ataavakpqavpnavaiae-----ivspitgvdvvalodqv 515  
Qy 325 SDAMPASGLSGVAIVPTKQQLVSPVSGKIVVAFPSGHAFVRTKAEDGSGNVDILMHIG 384  
Dy 516 pdeafaskavgdvavkptdkivvpaagtivkifntnhafclele----kgaieivvhmg 571  
Qy 385 FDTVNLNGTHFNPLKKQDEVKAGELLCEDFIDAIKAAGYEVVTPPIVVSNNYKKTGPVNTY 444  
Dy 572 idtvaldegkfriveegavqsagqpillemdldylnanarismispvvcnsiddfsgliik 631  
Qy 445 GLGEIEAGANLLNVAKK 461  
Dy 632 agghivagqtplyeikk 648

RESULT 8  
AAB66723  
ID AAB66723 standard; protein; 135 AA.

AC AAB66723;  
XX 09-APR-2001 (first entry)  
XX C.glutamicum phosphoenolpyruvate protein #7.  
XX Phosphoenolpyruvate; sugar phosphotransferase system; PTS.  
XX Corynebacterium glutamicum.

XX WO200102583-A2.  
XX 11-JAN-2001.  
XX 27-JUN-2000; 2000WO-IB00973.  
XX 01-JUL-1999; 99US-0142691.  
XX 23-AUG-1999; 99US-0150310.  
XX 03-SEP-1999; 99DE-1042095.  
XX 03-SEP-1999; 99DE-1042097.  
XX (BADI ) BASF AG.  
XX Pompejus M, Kroeger B, Schroeder H, Zeider O, Haberhauer G;  
XX WPI; 2001-080989/09.

XX Corynebacterium glutamicum nucleic acids encoding phosphoenolpyruvate:  
XX sugar phosphotransferase system proteins or their portions, useful for  
XX typing or identifying C. glutamicum or related bacteria, and as markers  
XX for transformation -  
XX Claim 36; Page 142-143; 144pp; English.  
XX The present invention relates to Corynebacterium glutamicum  
XX phosphoenolpyruvate: sugar phosphotransferase system (PTS) proteins.  
XX The PTS nucleic acids and proteins are useful in the  
XX identification of microorganisms which can be used to produce fine  
XX chemicals, for modulating fine chemical production in C. glutamicum or  
XX related bacteria, the typing or identification of C. glutamicum or  
XX related bacteria, as reference points for mapping C. glutamicum genome,  
XX and as markers for transformation.

XX Sequence 135 AA;

Query Match 9.9%; Score 234; DB 22; Length 135;  
Best Local Similarity 37.5%; Pred. No. 1.7e-15;  
Matches 51; Conservative 25; Mismatches 56; Indels 4; Gaps 1;

Qy 326 DAMFASGLSGVAIVPTKQQLVSPVSGKIVVAFPSGHAFVRTKAEDGSGNVDILMHIGF 385  
Dy 2 dpifaagklpgiaiqptgntvpadatvilvqkshavairl----dsgveilvhvql 57  
Qy 386 DTVNLNGTHFNPLKKQDEVKAGELLCEDFIDAIKAAGYEVVTPPIVVSNNYKKTGPVNTY 445  
Dy 58 dtvqlggsegftvhvrrgqvkagdpilifdadfirskdlpitpvvvsnaakfgelegip 117  
Qy 446 LGEIEAGANLLNVAKK 461  
Dy 118 adqanssttvikvngk 133

RESULT 9  
AAY20012  
ID AAY20012 standard; Protein; 167 AA.

XX AAY20012;  
XX 19-JUL-1999 (first entry)  
XX B. burgdorferi antigenic protein, f346.aa.  
XX Antigenic protein; vaccine; Lyme disease; infection; detection.  
XX Borrelia burgdorferi.

XX WO9859071-A1.  
XX 30-DEC-1998.  
XX 18-JUN-1998; 98WO-US12718.  
XX 03-SEP-1997; 97US-0057483.  
XX 20-JUN-1997; 97US-0050359.  
XX 22-JUL-1997; 97US-0053344.  
XX 22-JUL-1997; 97US-0053377.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX (MEDI-) MEDIMUNE INC.

XX Choi GH, Erwin AL, Hanson MS, Lathigra R;  
XX WPI; 1999-189980/16.  
XX N-PSDB; AAX61709.

XX New isolated Borrelia burgdorferi nucleic acids - used to develop  
XX products for the diagnosis, prevention and treatment of diseases  
XX caused by Borrelia, particularly Lyme disease

XX Claim 12; Page 161; 275pp; English.

XX This sequence represents a Borrelia burgdorferi (Bb) protein of the  
XX invention, which is suitable for use in a vaccine. The Bb polypeptides  
XX can be used in vaccines for eliciting protective antibodies to members of  
XX the Borrelia genus, particularly for the use against Lyme disease in  
XX humans and animals. They can be used for preventing or attenuating an  
XX infection caused by a member of the Borrelia genus. The products can also  
XX be used for detection of members of the Borrelia genus.

XX Sequence 167 AA;

Query Match 9.9%; Score 233.5; DB 20; Length 167;  
Best Local Similarity 34.8%; Pred. No. 2.6e-15;  
Matches 46; Conservative 30; Mismatches 51; Indels 5; Gaps 2;

Db	3	fiitnlitairigevllidplignaidrtserwgfkwvvgvggisslallalftdfgg1	62
Qy	112	-----PVGGEL-LFGIAYSPYIV-----TGLHQSPFPIELE-----LFNOGGS-----	147
Db	63	qngnpvylvlfigivylimdfysfkdtgfwmipalsldsrerekstsfarvgstigan	122
Qy	148	-----PIETASMAN-----IAQGAACLAVFPLFLAKSEKLKLAGASGVSAVLGITP-	194
Db	123	lvgvwtptllffsaskanpogdkqg-----wffalivaivgi--ltsitvlg9thevk	175
Ov	195	-----AIFGVNLRLRWPFF-----IGIGTA-----AIGG	218

[illegible]

RESULT	11	
AAV20013		
ID	AAV20013 standard; Protein; 135 AA.	
XX		
XX	AAV20013;	
XX		
XX		
DT	19-JUL-1999 (first entry)	
XX		
DE	B. burgdorferi antigenic protein, t346.aa.	
XX		
KW	Antigenic protein; vaccine; Lyme disease; infection; detection.	
XX		
OS	Borrelia burgdorferi.	
XX		
PN	WO9859071-A1.	
XX		
PD	30-DEC-1998.	
XX		
PF	18-JUN-1998; 98WO-US12718.	
XX		
PR	03-SEP-1997; 97US-0057483.	
PR	20-JUN-1997; 97US-0050359.	
PR	22-JUL-1997; 97US-0053344.	
PR	22-JUL-1997; 97US-0053377.	
XX		
PA	(HUMA-) HUMAN GENOME SCI INC.	

XX Choi GH, Erwin AL, Hanson MS, Lathigra R;  
XX  
XX  
XX  
DR WPI; 1999-199980/16.  
DR N-PSDB; AAX61710.  
XX  
XX New isolated *Borrelia burgdorferi* nucleic acids - used to develop





```

PF 31-MAR-2000; 2000WO-US08621.
XX
XX 31-MAR-1999; 99US-0127607.
XX 02-APR-1999; 99US-0127636.
PR 05-APR-1999; 99US-0127728.
XX 30-MAR-2000; 2000US-0540763.
PA (CURA-) CURAGEN CORP.
XX
XX Shinkets RA, Leach M;
XX PI
XX WPI; 2000-602362/57.
DR N-PSDB; AAC75000.
XX
XX Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease -
XX
XX Claim 11; Page 1038; 5507pp; English.
XX
XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytostatic; hepatotropic; vulnerrary;
CC osteoporotic; antiparkinsonian; nootropic; neuroprotective;
CC osteopathic; anticoagulant; antiarthritic; immunosuppressant;
CC immunomutulant; cardiant; thrombolytic; coagulant; vasotropic;
CC antiidiabetic; hypotensive; dermatological; immunosuppressive;
CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC antithyroid; and antianaemic. The sequences can be used for determining
CC the presence of or predisposition to, or preventing or treating
CC pathological conditions associated with an ORFX-associated disorder. The
CC nucleic acids can be used to express ORFX proteins in gene therapy
CC vectors. The proteins and nucleic acids may be used to treat cancers,
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC graft vs host disease, cardiovascular disease, diabetes mellitus,
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.
XX
XX Sequence 108 AA;
SQ
Query Match 6.5%; Score 154; DB 21; Length 108;
Best Local Similarity 32.7%; Pred. No. 1.2e-07;
Matches 32; Conservative 19; Mismatches 43; Indels 4; Gaps 1;
QY 300 PAEFNSDSTIIQAPLTEALTSVSDAMPASGKLGSGVAIVPTKGOLSPVSQGIWVAF 359
Db : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
14 psslkpsllkivaplggillvpldqvpdvfaqkmvgdglslpdsnellapvaqtqlh 73
QY 360 PSGHAFAYRTKAEDGSNVDILMHIGFTVNLNGTHFPNP 397
Db : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
74 nahaittcp----egievhlghldtvmrlrgdsvyp 107

```

Search completed: March 21, 2002, 16:22:45  
Job time: 221 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 21, 2002, 16:23:20 ; Search time 17.08 seconds  
(without alignments)  
1004.633 Million cell updates/sec

Title: US-09-604-231-2  
Perfect score: 2363  
Sequence: 1 MANVFPISLVNGYDVAAATMAA.....IEAGANLLNVAKKEAVPATP 468

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_39;\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1058.5	44.8	651	1	PTSA_PEDPE
2	953.5	40.4	664	1	PTSA_STRMU
3	701	29.7	480	1	PTSB_STAXY
4	678	28.7	474	1	PTSB_PASMU
5	657	27.8	609	1	PTBA_BACSU
6	646	27.3	479	1	PTSB_VIBAL
7	640	27.1	631	1	PTBA_ERWCH
8	570	24.1	470	1	PTTB_BACSU
9	567.5	24.0	625	1	PTBA_ECOLI
10	492	20.8	456	1	PTSB_SALTY
11	489	20.7	456	1	PTSB_KLEPN
12	462.5	19.6	674	1	PTSA_CORGL
13	406	17.2	459	1	SACK_BACSU
14	390	16.5	460	1	PTSB_BACSU
15	367	15.5	473	1	PTTB_ECOLI
16	321.5	13.6	699	1	PTGA_BACSU
17	317.5	13.4	651	1	PTAA_KLEPN
18	298.5	12.6	648	1	PTAA_ECOLI
19	294.5	12.5	324	1	PTGA_BACST
20	287.5	12.2	631	1	YBFS_BACSU
21	270.5	11.4	485	1	PTDA_ECOLI
22	256	10.8	189	1	PTGA_BORBU
23	255	10.8	168	1	YPOE_BACSU
24	254.5	10.8	634	1	LACY_STRTR
25	250.5	10.6	474	1	YFEV_ECOLI
26	249.5	10.6	154	1	PTGA_MYCCA
27	246	10.4	168	1	PTGA_ECOLI
28	244.5	10.3	482	1	PTSB_VIBCH
29	244	10.3	168	1	PTGA_SALTY
30	238.5	10.1	165	1	PTGA_HAEIN
31	233	9.9	627	1	LACY_LACDE
32	231.5	9.8	161	1	PTGA_BUCAL
33	230.5	9.8	940	1	PTGA_MYCPN

34	221	9.4	641	1	RAFP_PEDPE	P43466 pediococcus
35	204	8.6	908	1	PTGA_MYCGE	P47315 mycoplasma
36	165.5	7.0	639	1	LACY_LEULA	O48624 leuconostoc
37	159	6.7	483	1	PTVB_ECOLI	P32154 escherichia
38	151.5	6.4	658	1	HRSA_ECOLI	P54745 escherichia
39	144.5	6.1	527	1	PTTB_BACSU	P54715 bacillus su
40	136.5	5.8	359	1	PTWC_ECOLI	P32672 escherichia
41	130	5.5	578	1	PTFB_RHOCA	P23387 rhodobacter
42	126.5	5.4	580	1	PTFB_XANCP	P23355 xanthomonas
43	123.5	5.2	632	1	PTMA_BUCAL	P57635 buchnera ap
44	122	5.2	694	1	PTFA_MYCPN	P75039 mycoplasma
45	120.5	5.1	577	1	PTLB_LACCA	P24400 lactobacill

ALIGNMENTS

RESULT	1	PTSA_PEDPE	STANDARD;	PRT;	651 AA.
ID	PTSA_PEDPE				
AC	P43470;				
DT	01-NOV-1995 (Rel. 32, Created)				
DT	01-NOV-1995 (Rel. 32, Last sequence update)				
DT	15-JUL-1999 (Rel. 38, Last annotation update)				
DE	PTS SYSTEM, SUCROSE-SPECIFIC IIABC COMPONENT (ETIABC-SCR) (SUCROSE-				
DE	PERMEASE IIABC COMPONENT) (PHOSPHOTRANSFERASE ENZYME II, ABC				
DE	COMPONENT) (EC 2.7.1.69) (EII-SCR).				
GN	SCRA.				
OS	Pediococcus pentosaceus.				
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;				
OC	Pediococcus.				
OX	NCBI_TaxID=1255;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=PEP1.0;				
RA	Leenhouts K.K.J., Bolhuis A.A., Kok J.J., Venema G.G.;				
RL	Submitted (XXX-1994) to the EMBL/GenBank/DBJ databases.				
CC	-!- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT				
CC	SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE				
CC	-TRANSPORT SYSTEM. THE IICD DOMAINS CONTAIN THE SUGAR BINDING SITE				
CC	AND THE TRANSMEMBRANE CHANNEL; THE IIA DOMAIN CONTAINS THE PRIMARY				
CC	PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HPR); IIA TRANSFERS ITS				
CC	PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO				
CC	THE SUGAR.				
CC	-!- CATALYTIC ACTIVITY: PROTEIN N-PHOSPHOISTIDINE + SUGAR =				
CC	PROTEIN HISTIDINE + SUGAR PHOSPHATE.				
CC	-!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.				
CC	-!- SIMILARITY: CONTAINS A PTS EIIA DOMAIN.				
CC	-!- SIMILARITY: CONTAINS A PTS EIIB DOMAIN.				
CC	-!- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>				
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).				
CC	-----				
DR	EMBL; Z32771; CAA83668.1; -				
DR	EMBL; L32093; AAA25567.1; -				
DR	HSSP; P20166; IAX3.				
DR	InterPro; IPR001127; PTS_EIIA.				
DR	InterPro; IPR001996; PTS_EIIB.				
DR	InterPro; IPR003352; PTS_EIIC.				
DR	Pfam; PF00358; PTS_EIIA.1; 1.				
DR	Pfam; PF00367; PTS_EIIB.1.				
DR	Pfam; PF02378; PTS_EIIC.1.				
DR	ProDom; PD001476; PTS_EIIB.1.				
DR	ProDom; PD002243; PTS_EIIA.1.				
DR	PROSITE; PS00371; PTS_EIIA.1; 1.				
DR	PROSITE; PS01035; PTS_EIIB.CYS; 1.				
KW	Phosphotransferase system; Sugar transport; Transferase;				

```
KW Phosphorylation; Transmembrane.
FT DOMAIN 1 40 EIIB DOMAIN.
FT DOMAIN ? ? EIIC DOMAIN.
FT DOMAIN 510 651 EIIA DOMAIN.
FT MOD_RES 25 25 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 324 324 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 562 562 PHOSPHORYLATION (BY SIMILARITY).
SQ SEQUENCE 651 AA; 68454 MW; C87BA09D550A77F8 CRC64;

Query Match 44.8%; Score 1058.5; DB 1; Length 651;
Best Local Similarity 46.7%; Pred. No. 1.2e-62;
Matches 221; Conservative 78; Mismatches 151; Indels 23; Gaps 5;

QY 1 MAMVPSLVNGYDYAATMAAGEMPMSLFGLDVAQAGYQGVTLVPLVSVWILATIEKFLH 60
DB 194 MIMVPLSLVNGYSVATMAAGKQVYVNVFGLHVAQAGYQGVTLVPLVGVAFILATLEKFFH 253
QY 61 KRLKGTADELITPVLTLLLTGFTFIAGIPAMRWVGDVLAHGLQGLYDFGPGVGLLFLGL 120
DB 254 KHKGAFDEFTTPEFAIVITGFLFTTIVGPVLRVTSDBALTNGLVLYNSTGWMGIFGL 313
QY 121 VYSIVITGLHQSPPELELF-----NQGSFIFATASMANIAGAACLAVFFLAKEKL 176
DB 314 LYSIVITGLHQTPPAETQLLANVAKTGGSFIPFVASMANIGQGAATLAIFATKSKQK 373
QY 177 KGLAGASGVSAVLGITPAIFGVNLRWRPFFIGTAAGIAGGALIALFNKAVAGAGF 236
DB 374 KALTSSAGVSALLGITPAIFGVNLRWKPFVFAIASGIASAFGLFHLVLSVAMGPSV 433
QY 237 LGVVSIDAPDMVFLVCAVVTFFTAAGAAIAYGLYLVRRNGSIDPDATAAPVPAGTKAE 296
DB 434 IGFISIASKSPIAPMLSAVSEVFAFTFIY----AKRTLGDRDQVKSPAPTSTV--- 486
QY 297 AEAPAEFNSDTIIQAPLTGEAIALSVSDAMFASGKLGSGVAIVPTKGOLVSPVSGKIV 356
DB 487 -----INVDEILISAPVTGASESLKQVNDQVFSAEIMGKGAALVPSDDQVAPADGVIT 540
QY 357 VAPPSGHAFVTRKAEDGSNDVILMHIGFDTVNLNGTHFNPLKQGDDEVKAGELLCEFDI 416
DB 541 VYDSHAYGKTKTA-----GAELIHLGLDVLNLSGHEFTTNVQKGDVTHQGDLLGTFDI 596
QY 417 DAIRKAGYEVPTTVVSNKKTGPVNTYGLGEIEAGANL--LNVAKKEAVPAT 467
DB 597 AALKANYPDPTVMLIVTNTANYANVERLKVNTVQAGEQLVALTAPASSVAAT 649

RESULT 2
ID PTSA_STRMU STANDARD; PRT; 664 AA.
AC P12655;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE PTS SYSTEM, SUCROSE-SPECIFIC IIABC COMPONENT (EIIBC-SCR) (SUCROSE-
DE PERMEASE IIABC COMPONENT) (PHOSPHOTRANSFERASE ENZYME II, ABC
DE COMPONENT) (EC 2.7.1.69) (EII-SCR).
GN SCRA.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GS-5;
RX MEDLINE=89123027; PubMed=2536656;
RA Sato Y., Poy F., Jacobson G.R., Kuramitsu H.K.;
RT "Characterization and sequence analysis of the scrA gene encoding
RT enzyme IISCR of the Streptococcus mutans
RT phosphoenolpyruvate-dependent sucrose phosphotransferase system.";
RL J. Bacteriol. 171:263-271(1989).
RN [2]
RP SEQUENCE OF 639-664 FROM N.A.
```

```
RC STRAIN=GS-5;
RX MEDLINE=93329360; PubMed=8336109;
RA Sato Y., Yamamoto Y., Kizaki H., Kuramitsu H.K.;
RT "Isolation, characterization and sequence analysis of the scrK gene
RT encoding fructokinase of Streptococcus mutans.";
RL J. Gen. Microbiol. 139:921-927(1993).
CC -I- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT
CC SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS). A MAJOR CARBOHYDRATE ACTIVE
CC -TRANSPORT SYSTEM. THE IICD DOMAINS CONTAIN THE SUGAR BINDING SITE
CC AND THE TRANSMEMBRANE CHANNEL; THE IIA DOMAIN CONTAINS THE PRIMARY
CC PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HPR); IIA TRANSFERS ITS
CC PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO
CC THE SUGAR.
CC -I- CATALYTIC ACTIVITY: PROTEIN N-PHOSPHOTRANSFERASE + SUGAR =
CC PROTEIN HISTIDINE + SUGAR PHOSPHATE.
CC -I- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -I- SIMILARITY: CONTAINS A PTS EIIB DOMAIN.
CC -I- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M22711; AAA26971.1; -.
DR EMBL; D13175; BAA02466.1; -.
DR PIR; B32243; B32243.
DR HSP; P08837; 2F3G.
DR InterPro; IPR001127; PTS_EIIB.
DR InterPro; IPR001996; PTS_EIIC.
DR InterPro; IPR003352; PTS_EIIC.
DR Pfam; PF00358; PTS_EIIB_1; 1.
DR Pfam; PF00367; PTS_EIIB; 1.
DR Pfam; PF02378; PTS_EIIC; 1.
DR ProDom; PD001476; PTS_EIIB; 1.
DR ProDom; PD002243; PTS_EIIB; 1.
DR PROSITE; PS00371; PTS_EIIB_1; 1.
DR PROSITE; PS01035; PTS_EIIB_CYS; 1.
DR Phosphotransferase system; Sugar transport; Transferase;
KW Phosphorylation; Transmembrane.
FT DOMAIN 1 40 EIIB DOMAIN.
FT DOMAIN ? ? EIIC DOMAIN.
FT MOD_RES 25 25 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 331 331 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 585 585 PHOSPHORYLATION (BY SIMILARITY).
SQ SEQUENCE 664 AA; 69988 MW; 809E63E32281A9A1 CRC64;

Query Match 40.4%; Score 953.5; DB 1; Length 664;
Best Local Similarity 42.6%; Pred. No. 1e-55;
Matches 205; Conservative 89; Mismatches 156; Indels 31; Gaps 9;

QY 1 MAMVPSLVNGYDYAATMAAGEMP-----MMSLFGLDVAQAGYQGVTLVPLVSVWILA 53
DB 194 MIMVAPGAANIIGLAANAPISKAATIGAYTGFWNIFGLHVTQASYTYQVIVPLVAVWLLS 253
QY 54 TIEKFLHRLKGTADFLITPVLTLLLTGFTTETAGPAMRWVGDVLAHGLQGLYDFGPGV 113
DB 254 ILEKFFHRLPSAVDFTTPLLVIITGFTTFFIVGPMKEVSDWLTNGIWLVDITGFL 313
QY 114 GLLFLGLVYSPVITGLHQSPPELELF---FNQG---GSFIFATASMANIAGAACLAV 167
DB 314 GMVFGALYSPVVTGLHQSPFALETQLISAFQNGTGHGDFIFVTASMANVAGAAATFAI 373
QY 168 FFAKSEKLLKGLAGASGVSAVLGITPAIFGVNLRWRPFFIGTAAGIAGGALIALFNK 227
DB 374 YFLTOKKMKGLSSSSGVSALLGITTEPALFGVNLKYRPFPPFCALIGSASAAITAGLLQV 433
```



QY 228 AVALGAGFLGVVSDIDAPDMVFMELVCAVVTETFAAGAAIYAGLYLVRRNGSIDPDATAP 287  
 DB 434 AVSLGAGFLGLFUSIRASSIPFVWCELLISFAIAFTVYGYG-----KTKAVDFRAEAA 488  
 QY 288 VPAGTTTAKAEAPAEFNSDST-----IIQAPLTGEALSSVSDAMFASGKLGSGVAI 340  
 DB 489 VEEAIEEVO-ETPEEAASAANKAQAQVTDVLAAPLAGEAVELTSTVDPFSEAMGKGAIAI 547  
 QY 341 VPTKGLVSPVSGKIYVAPSPSHAFVTRKAEDGSNDVILMHIGFTVNLNGTHFNPLKK 400  
 DB 548 KPSGNTVYAPVDGTQVAFDTGHAYGI--KSDNGA--EILIHIGIDTVSMGKGFQKQV 603  
 QY 401 QGDEVKAGELLCFEDIDAIKAAGYEVYTPIVVSNYKKTGPVNTY-GLGETEAGANLLNVA 459  
 DB 604 ADOKIKKGDVLGTFDSKTAEGLDNTMTFIVTNTADYASVETLASSGIVAVGDSLLEVK 663  
 QY 460 K 460  
 DB 664 K 664

RESULT 3  
 PTSE\_STAXY  
 ID PTSE\_STAXY STANDARD; PRT; 480 AA.  
 AC P51184;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE PTS SYSTEM, SUCROSE-SPECIFIC IIBC COMPONENT (EIIBC-SCR) (SUCROSE-  
 DE PERWEASE IIBC COMPONENT) (PHOSPHOTRANSFERASE ENZYME II, BC COMPONENT)  
 DE (EC 2.7.1.69) (EII-SCR).  
 GN SCRA.  
 OS Staphylococcus xylosoyus.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 OC Bacillus/Staphylococcus group; Staphylococcus.  
 OX NCBI\_TaxID=1288;  
 RP [1]  
 RC SEQUENCE FROM N.A.  
 RC STRAIN=DSM 20267 / C2A;  
 RX MEDLINE=94049686; PubMed=82322209;  
 RA Wagner E., Goetz F., Brueckner R.;  
 RT "Cloning and characterization of the scrA gene encoding the sucrose-  
 RT specific Enzyme II of the phosphotransferase system from  
 RT Staphylococcus xylosoyus".  
 RL Mol. Gen. Genet. 241:33-41(1993).  
 CC -!- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT  
 CC SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE  
 CC -TRANSPORT SYSTEM. THE IICD DOMAINS CONTAIN THE SUGAR BINDING SITE  
 CC AND THE TRANSMEMBRANE CHANNEL; THE IIA DOMAIN CONTAINS THE PRIMARY  
 CC PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HPR); IIA TRANSFERS ITS  
 CC PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO  
 CC THE SUGAR.  
 CC -!- CATALYTIC ACTIVITY: PROTEIN N-PHOSPHOHIISTIDINE + SUGAR =  
 CC PROTEIN HISTIDINE + SUGAR PHOSPHATE.  
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE.  
 CC -!- SIMILARITY: CONTAINS A PTS EIIIB DOMAIN.  
 CC -!- SIMILARITY: CONTAINS A PTS EIIIC DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X69800; CAA49461.1;  
 DR InterPro: IPR001996; PTS\_EIIB.  
 DR InterPro: IPR003352; PTS\_EIIC.  
 DR Pfam: PF00367; PTS\_EIIB; 1.  
 DR Pfam: PF02378; PTS\_EIIC; 1.  
 DR ProDom: PD001476; PTS\_EIIB; 1.  
 DR PROSITE: PS01035; PTS\_EIIB\_CYS; 1.

KW Phosphotransferase system; Sugar transport; Transferase;  
 KW Transmembrane; Inner membrane; Phosphorylation.  
 FT DOMAIN 1 ? EIIIB DOMAIN.  
 FT MOD\_RES 26 26 PHOSPHORYLATION (BY SIMILARITY).  
 FT MOD\_RES 325 325 PHOSPHORYLATION (BY SIMILARITY).  
 SQ SEQUENCE 480 AA; 51326 MW; AB4E1D9785D84E47 CRC64;

Query Match 29.7%; Score 701; DB 1; Length 480;  
 Best Local Similarity 51.4%; Pred. No. 3e-39;  
 Matches 146; Conservative 49; Mismatches 81; Indels 8; Gaps 4;

QY 1 MANVPSLVNGYDVAATMAAG-EMPWMSLEGLDVAAGYQGTVPVLVYVSWILATIEKPL 59  
 DB 194 MILVHPFLMSAYDYPKALEAGKEIPHNLFGLEINQVGGVQGLPMLVAVYIATIEKGL 253  
 QY 60 HKRLKGTADFLITPVLTLTLTGFTLFIAGPAMRWGDVLAHGLQGLYDFGPGVGLLFG 119  
 DB 254 RKVIPVLDNLLTPLLAILSTGFTSFVGLTRTGLYWLSDGLTWLYEFGGAIGLFG 313  
 QY 120 LVYSPVITGLHOSFPPIELELF----NOGGSFIFATASMANIAQAACLAFFLAK-SE 174  
 DB 314 LLYAPIVITGMHHSFIAETQLIADSSSTGSGSFIFPIATMSNIAQAALAAFFIIEK 373  
 QY 175 KLKGLAGASGVSAVLGITPEPAITGVNLRWPFEGITGAAGALIALFNKAVAGAA 234  
 DB 374 KLKGVASAGVSALLGITPEPAITGVNLRWPFEGITGAAGALIALFNKAVAGAA 433  
 QY 235 GLFGVVISIDAPD--WVMEFLVCAVVTETFAAGAAIYAGLYLVRRN 276  
 DB 434 GIPGFISIQNNQNLHYGIAMIIAFVAGVYVALSYRKKYRN 477

RESULT 4  
 PTSE\_PASMU  
 ID PTSE\_PASMU STANDARD; PRT; 474 AA.  
 AC Q9CJ22;  
 DT 20-AUG-2001 (Rel. 40, Created)  
 DT 20-AUG-2001 (Rel. 40, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE PTS SYSTEM, SUCROSE-SPECIFIC IIBC COMPONENT (EIIBC-SCR) (SUCROSE-  
 DE PERWEASE IIBC COMPONENT) (PHOSPHOTRANSFERASE ENZYME II, BC COMPONENT)  
 DE (EC 2.7.1.69) (EII-SCR).  
 GN SCRA OR PTSE OR PM1846.  
 OS Pasteurella multocida.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
 OC Pasteurella.  
 OX NCBI\_TaxID=747;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=PM70;  
 RC MEDLINE=21145866; PubMed=11248100;  
 RA May B.J., Zhang Q., Li L., Paustian M.L., Whittam T.S., Kapur V.;  
 RT "Complete genomic sequence of Pasteurella multocida Pm70".;  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).  
 CC -!- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT  
 CC SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE  
 CC -TRANSPORT SYSTEM. THE IICD DOMAINS CONTAIN THE SUGAR BINDING SITE  
 CC AND THE TRANSMEMBRANE CHANNEL; THE IIA DOMAIN CONTAINS THE PRIMARY  
 CC PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HPR); IIA TRANSFERS ITS  
 CC PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO  
 CC THE SUGAR (BY SIMILARITY).  
 CC -!- CATALYTIC ACTIVITY: PROTEIN N-PHOSPHOHIISTIDINE + SUGAR =  
 CC PROTEIN HISTIDINE + SUGAR PHOSPHATE.  
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE  
 CC (BY SIMILARITY).  
 CC -!- SIMILARITY: CONTAINS A PTS EIIIB DOMAIN.  
 CC -!- SIMILARITY: CONTAINS A PTS EIIIC DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC	EMBL; AE006222; AAK03930.1; -.	
DR	InterPro; IPR001996; PTS_EIIB.	
DR	InterPro; IPR003352; PTS_EIIG.	
DR	Pfam; PF00367; PTS_EIIB; 1.	
DR	Pfam; PF02378; PTS_EIIC; 1.	
DR	PROSITE; PS01035; PTS_EIIB_CVS; FALSE_NEG.	
KW	Phosphotransferase system; Transferase; Sugar transport; Plasmid;	
KW	Transmembrane; Inner membrane; Phosphorylation; Complete proteome.	
FT	DOMAIN 1 ? EIIB DOMAIN.	
FT	DOMAIN 109 474	
FT	TRANSMEM 109 129	POTENTIAL.
FT	TRANSMEM 158 178	POTENTIAL.
FT	TRANSMEM 182 202	POTENTIAL.
FT	TRANSMEM 229 249	POTENTIAL.
FT	TRANSMEM 264 284	POTENTIAL.
FT	TRANSMEM 303 323	POTENTIAL.
FT	TRANSMEM 345 365	POTENTIAL.
FT	TRANSMEM 376 396	POTENTIAL.
FT	TRANSMEM 403 423	POTENTIAL.
FT	TRANSMEM 444 464	POTENTIAL.
FT	MOD_RES 26 26	PHOSPHORYLATION (BY SIMILARITY).
FT	MOD_RES 324 324	PHOSPHORYLATION (BY SIMILARITY).
SQ	SEQUENCE 474 AA; 49908 MW; D26FA1A059603C01 CRC64;	

	EMBL; Z34526; CAA84286.1; -	
DR	EMBL; D31856; BAA06652.1; -	
DR	EMBL; D29985; BAA06256.1; -	
DR	EMBL; X85408; CAA59697.1; -	
DR	EMBL; Z99124; CAB15963.1; -	
DR	PIR; S47174; S47174.	
DR	HSSP; P20166; IAX3.	
DR	Subtilist; BG10934; bqlp.	
DR	InterPro; IPR001127; PTS_EIIA.	
DR	InterPro; IPR001996; PTS_EIIB.	
DR	InterPro; IPR003352; PTS_EIIC.	
DR	Pfam; PF00358; PTS_EIIA.1; 1.	
DR	Pfam; PF00367; PTS_EIIB; 1.	
DR	Pfam; PF02378; PTS_EIIC; 1.	
DR	ProDom; PD001476; PTS_EIIB; 1.	
DR	ProDom; PD002243; PTS_EIIA; 1.	
DR	PROSITE; PS00371; PTS_EIIA.1; 1.	
DR	PROSITE; PS01035; PTS_EIIB_CVS; 1.	
KW	Phosphotransferase system; Sugar transport; Transerase;	
KW	Phosphorylation; Transmembrane; Complete proteome.	
DOMAIN	1 41	EIIB DOMAIN.
FT	DOMAIN ?	EIIC DOMAIN.
FT	DOMAIN 480 609	EIIB DOMAIN.
FT	MOD_RES 25 25	PHOSPHORYLATION (BY SIMILARITY).
FT	MOD_RES 302 302	PHOSPHORYLATION (BY SIMILARITY).
FT	MOD_RES 532 532	PHOSPHORYLATION (BY SIMILARITY).
FT	TRANSMEM 112 132	POTENTIAL.
FT	TRANSMEM 141 161	POTENTIAL.



```

RESULT 7
PTBA_ERWCH STANDARD; PRT; 631 AA.
AC P26207;
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DE PTS SYSTEM, BETA-GLUCOSIDES-SPECIFIC IIBAC COMPONENT (EIIABC-BGL)
DE (BETA-GLUCOSIDES-PERMEASE IIBAC COMPONENT) (PHOSPHOTRANSFERASE
DE ENZYME II, ABC COMPONENT) (EC 2.7.1.69) (EII-BGL).
GN ARBF.
OS Erwinia chrysanthemi.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Pectobacterium.
OX NCBI_TaxID=556;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92121114; PubMed=1732212;
RA el Hassouni M., Henrissat B., Chippaux M., Barras F.;
RT "Nucleotide sequences of the arb genes, which control beta-glucoside
RT utilization in Erwinia chrysanthemi: comparison with the Escherichia
RT coli bgl operon and evidence for a new beta-glycosylase family
RT including enzymes from eubacteria, archaeobacteria, and humans.";
RL J. Bacteriol. 174:765-777(1992).
CC -1- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT
CC SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE
CC -TRANSPORT SYSTEM. THE IICD DOMAINS CONTAIN THE SUGAR BINDING SITE
CC AND THE TRANSMEMBRANE CHANNEL; THE IIA DOMAIN CONTAINS THE PRIMARY
CC PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HPR); IIA TRANSFERS ITS
CC PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO
CC THE SUGAR.
CC -1- FUNCTION: ACTS AS BOTH A KINASE AND A PHOSPHATASE ON ARBG (BY
CC SIMILARITY).
CC -1- CATALYTIC ACTIVITY: PROTEIN N-PHOSPHOHISTIDINE + SUGAR =
CC PROTEIN HISTIDINE + SUGAR PHOSPHATE.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE.
CC -1- SIMILARITY: CONTAINS A PTS EIIA DOMAIN.
CC -1- SIMILARITY: CONTAINS A PTS EIIB DOMAIN.
CC -1- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M81772; AAA24814.1; -.
CC PIR; B42603; B42603.
CC HSSP; P20166; 1AX3.
CC InterPro; IPR001127; PTS_EIIA.
CC InterPro; IPR001996; PTS_EIIB.
CC InterPro; IPR003352; PTS_EIIC.
CC Pfam; PF00358; PTS_EIIA.1; 1.
CC Pfam; PF00367; PTS_EIIB.1.
CC Pfam; PF02378; PTS_EIIC.1.
CC ProDom; PD001476; PTS_EIIB; 1.
CC ProDom; PD002243; PTS_EIIA; 1.
CC PROSITE; PS00371; PTS_EIIA.1; 1.
CC PROSITE; PS01035; PTS_EIIB.CVS; 1.
CC Phosphotransferase system; Sugar transport; Transferase;
CC Phosphorylation; Transmembrane; Inner membrane.
CC DOMAIN 1 41 EIIB DOMAIN.
CC DOMAIN ? ? EIIC DOMAIN.
CC DOMAIN 501 631 EIIA DOMAIN.
CC MOD_RES 26 26 PHOSPHORYLATION (BY SIMILARITY).
CC MOD_RES 309 309 PHOSPHORYLATION (BY SIMILARITY).
CC MOD_RES 553 553 PHOSPHORYLATION (BY SIMILARITY).
CC TRANSMEM 50 70 POTENTIAL.
CC TRANSMEM 99 119

```

```

FT TRANSMEM 120 140 POTENTIAL.
FT TRANSMEM 146 166 POTENTIAL.
FT TRANSMEM 175 195 POTENTIAL.
FT TRANSMEM 206 226 POTENTIAL.
FT TRANSMEM 248 268 POTENTIAL.
FT TRANSMEM 272 292 POTENTIAL.
FT TRANSMEM 295 315 POTENTIAL.
FT TRANSMEM 328 348 POTENTIAL.
FT TRANSMEM 358 378 POTENTIAL.
FT TRANSMEM 385 405 POTENTIAL.
FT TRANSMEM 407 427 POTENTIAL.
FT TRANSMEM 434 454 POTENTIAL.
SQ SEQUENCE 631 AA; 66984 MW; 7D0BD27A36BFFDFD CRC64;

Query Match 27.1%; Score 640; DB 1; Length 631;
Best Local Similarity 33.8%; Pred. No. 4.2e-35;
Matches 154; Conservative 86; Mismatches 200; Indels 16; Gaps 8;

QY 3 MVPSLVNGYDVAAATMAAGMPWMSLFLGLDVAQAGYOGTVLPVLVLSWILATIEKFLHR 62
DB LVHPSMTAATN---AMQAPDSTLHFGIPITFTINSSSVIPILFASVSKLEKPLNRW 240
QY 63 LKGTADFLITPVLTLLTFTIATGPAWRWGDVLAHGLQGLYDFGPGVGLLFLGVY 122
DB LHANIRNFFTPLLCIVISVPLTFLIGPSATWLSQMLAGGVQWLYGLNSLLAGAVMGALW 300
QY 123 SPIVITGLHOSFPPIELFNQ--GGSFIFATASMANIAQAACLAFLAKSEKLGAG 181
DB 301 QVCVIFGLHGFVPLMLNFFSVIGHDTLLPLLVPAVLGQAGATLGLVLLRTQDLKRKG 360
QY 182 ASGVSALVGTGTEPAIFGVNLRWPFPIGTGTAAGGALIALFNKAVALCAAGFLGVVS 241
DB 361 SPSAAIFGTEPAVGVTLPLRPFFGCGIGALGAARVMGYAHTTWSFGFPISFTQ 420
QY 242 IDAPDMVNLV-CAVVTFFIAG-AAIAYGLYLVRRNGSIDPDATAAPVPAGTTKABEA 299
DB 421 VIPPTGVDSSWAIVIGTLLAFAPALTSWSFGVPKD---ETQPAADSPA--VLAETQA 475
QY 300 PAEFSNDSTIIQAPLTGEATLSSVDAMFASGLSGVAIVPTKGLVSPVSKIVVAF 359
DB 476 NAGAVRDETIF-SPLAGEVLLLEQVADRTFASGVNMGKGIARPTQGRLYAPVDGTVASLF 534
QY 360 PSGHAFAVRTKAEDGSNVVDLMHIGFDTVNLNGTHFPLKKGQDEVKAGELCCFDDAI 419
DB 535 KTHHAIGLASR---GGAELVLIHVIDTVLDGRYFTPHVRVGVDRVQGGDLLLEDFGP 590
QY 420 KAAGYEVTTPIVSNYKKTGPVNTYGLGEIEAGANL 455
DB 591 EAAGYDLTTPITVITNSDYRGVSPVASKVDANAPL 626

```

```

RESULT 8
PTTB_BACSU STANDARD; PRT; 470 AA.
ID PTTB_BACSU
AC P39794; O34771;
DT 01-FEB-1995 (Rel. 31, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROBABLE PTS SYSTEM, TREHALOSE-SPECIFIC IIBC COMPONENT (EIIBC-TRE)
DE (TREHALOSE-PERMEASE IIBC COMPONENT) (PHOSPHOTRANSFERASE ENZYME II, BC
DE COMPONENT) (EC 2.7.1.69) (EII-TRE).
GN TREP OR TREP.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168 / MABURG;
RX MEDLINE=97074649; PubMed=8917076;
RA Schoeck F., Dahl M.K.;
RT "Analysis of DNA flanking the treA gene of Bacillus subtilis reveals

```

genes encoding a putative specific enzyme IITre and a potential regulator of the trehalose operon.";

Gene 175:59-63(1996).

[2]

SEQUENCE FROM N.A.

STRAIN=AC327;

MEDLINE=97124190; PubMed=8969503;

Yamamoto H., Uchiyama S., Sekiguchi J.;

"Cloning and sequencing of a 40.6 kb segment in the 73 degrees-76

degrees region of the Bacillus subtilis chromosome containing genes

for trehalose metabolism and acetoin utilization.";

Microbiology 142:3057-3065(1996).

[3]

SEQUENCE FROM N.A.

STRAIN=AC327;

MEDLINE=97417486; PubMed=9272861;

Yamamoto H., Uchiyama S., Nugroho F.A., Sekiguchi J.;

"Cloning and sequencing of a 35.7 kb in the 70 degree-73 degree region

of the Bacillus subtilis genome reveal genes for a new two-component

system, three spore germination proteins, an iron uptake system and a

general stress response protein.";

Gene 194:191-199(1997).

[4]

SEQUENCE OF 324-470 FROM N.A.

STRAIN=168;

MEDLINE=95379486; PubMed=7651129;

Helfert C., Gotsche S., Dahl M.K.;

"Cleavage of trehalose-phosphate in Bacillus subtilis is catalysed by

a phospho-alpha-(1-1)-glucosidase encoded by the treA gene.";

Mol. Microbiol. 16:111-120(1995).

-1- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT

SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE

-TRANSPORT SYSTEM. THE IICD DOMAINS CONTAIN THE SUGAR BINDING SITE

AND THE TRANSMEMBRANE CHANNEL; THE IIA DOMAIN CONTAINS THE PRIMARY

PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HPR); IIA TRANSFERS ITS

PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO

THE SUGAR.

-1- CATALYTIC ACTIVITY: PROTEIN N-PHOSPHOTRANSFERASE + SUGAR -

PROTEIN HISTIDINE + SUGAR PHOSPHATE.

-1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE.

-1- SIMILARITY: CONTAINS A PTS EIIB DOMAIN.

-1- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.

-----

This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL Outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

EMBL; 254245; CAA91014.1; -

DR EMBL; D83967; BAA23409.1; -

DR EMBL; D86417; BAA22289.1; -

DR EMBL; X80203; CAA56494.1; -

DR EMBL; 299108; CAB12609.1; -

DR HSP; P05053; IIBA.

DR Subtilist; BG11009; trep.

DR InterPro; IPR001996; PTS\_EIIB.

DR InterPro; IPR003352; PTS\_EIIC.

DR Pfam; PF00367; PTS\_EIIB; 1.

DR Pfam; PF02378; PTS\_EIIC; 1.

DR ProDom; PD001476; PTS\_EIIB; 1.

DR PROSITE; PS01035; PTS\_EIIB\_CYS; 1.

KW Phosphotransferase system; Sugar transport; Transferase;

Transmembrane; Inner membrane; Phosphorylation; Complete proteome.

FT DOMAIN 1 ?

FT MOD\_RES 26 26

FT TRANSMEM 110 130

FT TRANSMEM 160 180

FT TRANSMEM 183 203

FT TRANSMEM 234 254

FT TRANSMEM 234 254

FT TRANSMEM 234 254

FT TRANSMEM 234 254

FT TRANSMEM 234 254

FT TRANSMEM 234 254

FT TRANSMEM 234 254

FT TRANSMEM 234 254

FT TRANSMEM 234 254

FT TRANSMEM 234 254

FT TRANSMEM 234 254

FT TRANSMEM 234 254

FT TRANSMEM 234 254

FT TRANSMEM 234 254

FT TRANSMEM 234 254

FT TRANSMEM 234 254

FT TRANSMEM 234 254

FT TRANSMEM 234 254

FT TRANSMEM 234 254

FT TRANSMEM 263 283 POTENTIAL.

FT TRANSMEM 301 321 POTENTIAL.

FT TRANSMEM 326 346 POTENTIAL.

FT TRANSMEM 347 367 POTENTIAL.

FT TRANSMEM 375 395 POTENTIAL.

FT TRANSMEM 403 423 POTENTIAL.

FT TRANSMEM 443 463 POTENTIAL.

FT CONFLICT 140 140 F -> S (IN REF. 1).

FT CONFLICT 363 363 M -> L (IN REF. 1 AND 4).

FT CONFLICT 465 465 A -> G (IN REF. 1 AND 4).

SQ SEQUENCE 470 AA; 49999 MW; 7A741850A2697D53 CRC64;

Query Match 24.1%; Score 570; DB 1; Length 470;

Best Local Similarity 42.8%; Pred. No. 1.2e-30;

Matches 116; Conservative 54; Mismatches 99; Indels 2; Gaps 1;

QY 3 MFPPSLVNGYDVAAATMAAGEMPMSLFLGIDVAQAQYQGTVPVLVSVILATIERKFLHKKR 62

DB 197 LVHPDLLNANGVGAERQSGEIPVNLFLGVKQVGVQGVLPILLASYMLAKIEVFLTKR 256

QY 63 LKGTADFLLITPVLTLTLTGFTTFAIGPAMRWGVDLAHGLQGLYDFGPGVGLLFLGLVY 122

DB 257 TPEGIQLLVAVPITLLTGTGFASFIIGPITFAIGNVLTSLISVFGSFAALGGLLYGGFY 316

QY 123 SPIVITGLHQSPFPPIELEFNQ--GGSFIFATASMANIAQAACLAFLVFLAKSEKLGCLA 180

DB 317 SALVTGTHHTFLAVLDLQIGSKLGTFLWPMALSNIAQSAALAMMFIVKDEKOKGLS 376

QY 181 GASGVSAVLGITEPAIFGVNLRWLPFFIGITAAIGGALIALFNKIKAVAGAAAGFLGVV 240

DB 377 LTSGLSIVLGITTEPAIFGVNLRWLPFFIGITAAIGGALIALFNKIKAVAGAAAGFLGVV 240

QY 241 SIDADPMVNLVCAVVTFFIAFGAAIAYGLY 271

DB 437 SIMSQYGAFAIGMAIVLIVPEAGTVAYARE 467

RESULT 9

PTBA\_ECOLI

ID PTBA\_ECOLI STANDARD; PRT; 625 AA.

AC P08722;

DT 01-AUG-1988 (Rel. 08, Created)

DT 01-AUG-1988 (Rel. 08, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE PTS SYSTEM, BETA-GLUCOSIDES-SPECIFIC IIABC COMPONENT (EIIABC-BGL)

DE (BETA-GLUCOSIDES-PERMEASE IIABC COMPONENT) (PHOSPHOTRANSFERASE

ENZYME II, ABC COMPONENT) (EC 2.7.1.69) (EII-BGL).

GN BGLF OR BGLC OR BGLS OR B3722.

OS Escherichia coli.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Escherichia.

OX NCBI\_TaxID=562;

[1]

RN SEQUENCE FROM N.A.

RC STRAIN=K12;

RX MEDLINE=88009877; PubMed=3309161;

RA Bramley H.F., Kornberg H.L.;

"Nucleotide sequence of bgIC, the gene specifying enzymebgl of the

PEP-sugar phosphotransferase system in Escherichia coli K12, and

overexpression of the gene product";

RL J. Gen. Microbiol. 133:563-573(1987).

[2]

RN SEQUENCE FROM N.A.

RC STRAIN=K12;

RX MEDLINE=87222180; PubMed=3034860;

RA Schnetz K., Toloczkyi C., Rak B.;

"Beta-glucoside (bgl) operon of Escherichia coli K-12: nucleotide

sequence, genetic organization, and possible evolutionary

relationship to regulatory components of two Bacillus subtilis

genes.";

RL J. Bacteriol. 169:2579-2590(1987).

[3]

SEQUENCE FROM N.A.  
 STRAIN=K12 / MG1655;  
 MEDLINE=93315143; PubMed=7686882;  
 Burland V.D., Plunkett G. III, Daniels D.L., Blattner F.R.;  
 "DNA sequence and analysis of 136 kilobases of the *Escherichia coli*  
 genome: organizational symmetry around the origin of replication.";  
 RL Genomics 16:551-561(1993).  
 CC -1- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT  
 CC SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS). A MAJOR CARBOHYDRATE ACTIVE  
 CC -TRANSPORT SYSTEM. THE IIDC DOMAINS CONTAIN THE SUGAR BINDING SITE  
 CC AND THE TRANSMEMBRANE CHANNEL; THE IIA DOMAIN CONTAINS THE PRIMARY  
 CC PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HPR); IIA TRANSFERS ITS  
 CC PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO  
 CC THE SUGAR.  
 CC -1- FUNCTION: ACTS AS BOTH A KINASE AND A PHOSPHATASE ON BGLG.  
 CC -1- CATALYTIC ACTIVITY: PROTEIN N-PHOSPHOHISTIDINE + SUGAR =  
 CC PROTEIN HISTIDINE + SUGAR PHOSPHATE.  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE.  
 CC -1- SIMILARITY: CONTAINS A PTS EIIA DOMAIN.  
 CC -1- SIMILARITY: CONTAINS A PTS EIIIB DOMAIN.  
 CC -1- SIMILARITY: CONTAINS A PTS EIIIC DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M15746; AAA83837.1; -;  
 DR EMBL; M16487; AAA23510.1; -;  
 DR EMBL; L10328; AAA62073.1; -;  
 DR EMBL; AE000449; AAC76745.1; -;  
 DR PIR; C25977; C25977.  
 DR PIR; A47616; A47616.  
 DR HSP; P20166; 1AX3.  
 DR EcoGene; EG10115; bgfF.  
 DR InterPro; IPR001127; PTS\_EIIA.  
 DR InterPro; IPR001996; PTS\_EIIB.  
 DR InterPro; IPR003382; PTS\_EIIC.  
 DR Pfam; PF003358; PTS\_EIIA\_1; 1.  
 DR Pfam; PF003667; PTS\_EIIB; 1.  
 DR Pfam; PF02378; PTS\_EIIC; 1.  
 DR ProDom; PD001476; PTS\_EIIB; 1.  
 DR ProDom; PD002243; PTS\_EIIA; 1.  
 DR PROSITE; PS00371; PTS\_EIIA\_1; 1.  
 DR PROSITE; PS01035; PTS\_EIIB\_CYS; 1.  
 KW Phosphotransferase system; Sugar transport; Transferase;  
 KW Phosphorylation; Transmembrane; Inner membrane; Complete proteome.  
 FT DOMAIN 1 39 EIIB DOMAIN.  
 FT DOMAIN 2 625 EIIC DOMAIN.  
 FT DOMAIN 3 625 EIIA DOMAIN.  
 FT MOD\_RES 24 24 PHOSPHORYLATION (BY SIMILARITY).  
 FT MOD\_RES 306 306 PHOSPHORYLATION (BY SIMILARITY).  
 FT MOD\_RES 547 547 PHOSPHORYLATION (BY SIMILARITY).  
 FT TRANSMEM 59 79 POTENTIAL.  
 FT TRANSMEM 100 120 POTENTIAL.  
 FT TRANSMEM 141 161 POTENTIAL.  
 FT TRANSMEM 167 187 POTENTIAL.  
 FT TRANSMEM 203 223 POTENTIAL.  
 FT TRANSMEM 245 265 POTENTIAL.  
 FT TRANSMEM 285 305 POTENTIAL.  
 FT TRANSMEM 325 345 POTENTIAL.  
 FT TRANSMEM 355 375 POTENTIAL.  
 FT TRANSMEM 381 401 POTENTIAL.  
 FT TRANSMEM 404 424 POTENTIAL.  
 FT TRANSMEM 432 452 POTENTIAL.  
 FT TRANSMEM 490 510 POTENTIAL.  
 FT TRANSMEM 526 546 POTENTIAL.  
 FT TRANSMEM 625 AA; 66482 MW; 9E12A2207125C4E6 CRC64;  
 SQ SEQUENCE

Query Match 24.0%; Score 567.5; DB 1; Length 625;  
 Best Local Similarity 30.8%; Pred. No. 2.5e-30;  
 Matches 142; Conservative 91; Mismatches 211; Indels 17; Gaps 7;  
 QY 2 AMVPPSLVNGYDYAATMAAGEMPMSLFGDLDAQAQGVGVPLVPLVSVWILATEKEFLHK 61  
 DB 180 ALVHLPLTLTAFENGQKADALGL---DFLGIPVTLTLLNYSVSSVPIIFSALWCSILERRLNA 236  
 QY 62 RLKGTADFLTPVLTLLTGLTFIAIGPAMRWGDVLAHGLQGLYDFGGPVGGLLGLV 121  
 DB 237 WLPSAIKNFTPLCLMVTITPVTFLLVGLPLSTWISLAIAGYLWLXQAVPAFAGAVMGGF 296  
 QY 122 YSPITVITGLHQSPPIELFLFN-QGGSFIFATASMANIAQGAACLAFFFLAKSEKLGKLA 180  
 DB 297 WQIFVMEGLHWGLVPCINNFVTLGYDTMPLMPMAQVGAALGVFLCERDAQKQVVA 356  
 QY 181 GASGVSAVLGTEPAIFGVNLRWPFPIGIGTAAIGGALIALINIKAVAGALGAAGFLGVV 240  
 DB 357 GSAALTSFGITEPAFVGVNLRWPFPIGIGTAAIGGALIALINIKAVAGALGAAGFLGVV 416  
 QY 241 SIDAPDMVMEFLCAVVTFFTAAGAAIAYGLYVRRNSIDPD-ATAAPVPAGTTKAEAA 299  
 DB 417 QTPSTGIDFTWASV-----IGGVIAIGCAV---GTVMLHFTAKRQAQAPQKPTP 468  
 QY 300 PAEFSNDSTIIQAPLTGEAIALSVSDAMFASGKLGSGVAIVPTKGQVSPVSGKIYVAF 359  
 DB 469 EVITPPQGGICSPMTGCEIVPLIHVADTTFASGLLGKGIALLPSVGEVRSVPVAGRIASLF 528  
 QY 360 PSGHAFVAVRTKAEDGSNVDILMHIGFDTVNLGTHFNPLKKQGDVYKAGELLCEFDIDAI 419  
 DB 529 ATLHAIGI--ESDDG--VEILIHVGIDTVRLDKGFFSAHVNVGDKVNTGDRLLISFDIPAI 584  
 QY 420 KAAGYEVTTPIVSVNKKTPGVNTYGLGTEAGANLNVAK 460  
 DB 585 REAGFDLTPVLTNSDDFTDVLPHGTAQISAGEPLLSIIR 625  
 RESULT 10  
 ID PTSB\_SALTY STANDARD; PRT; 456 AA.  
 AC P08470;  
 DT 01-AUG-1988 (Rel. 08, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE PTS SYSTEM, SUCROSE-SPECIFIC IIBC COMPONENT (EIIBC-SCR) (SUCROSE-  
 DE PERMEASE IIBC COMPONENT) (PHOSPHOTRANSFERASE ENZYME II, BC COMPONENT)  
 DE (EC 2.7.1.69) (EII-SCR).  
 GN SCRA.  
 OS *Salmonella typhimurium*.  
 OG Plasmid pUR400.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC *Salmonella*  
 OC NCBI\_TaxID=602;  
 RN [1]  
 RX MEDLINE=88216186; PubMed=3285123;  
 RA Ebner R., Lengeler J.W.;  
 RT "DNA sequence of the gene *scrA* encoding the sucrose transport protein  
 RT EnzymeII(Scr) of the phosphotransferase system from enteric bacteria:  
 RT homology of the EnzymeII(Scr) and EnzymeII(Bgl) proteins.";  
 RL Mol. Microbiol. 2:9-17(1988).  
 RN [2]  
 RP REVISIONS.  
 RX MEDLINE=94018607; PubMed=8412665;  
 RA Jahreis K., Lengeler J.W.;  
 RT "Molecular analysis of two *ScrR* repressors and of a *ScrR*-*Frur* hybrid  
 RT repressor for sucrose and D-fructose specific regulons from enteric  
 RT bacteria.";  
 RL Mol. Microbiol. 9:195-209(1993).  
 RN [3]  
 RP SEQUENCE OF 1-7 FROM N.A.  
 RC STRAIN=6153-62;  
 RX MEDLINE=91100329; PubMed=1846143;



SQ SEQUENCE 456 AA; 48022 MW; 4AFDF5405CAEF66 CRC64;

Query Match 20.7%; Score 489; DB 1; Length 456;

Best Local Similarity 38.5%; Pred. No. 2.6e-25;

Matches 104; Conservative 51; Mismatches 101; Indels 14; Gaps 3;

QY 3 MVFPLSVNGYDVA--TMAAGEMPMSLFGDVAQAQYQCTVPLVSVSWILATIEKFL 59  
DB 186 LTHPALTNAWGVAAGFHT-----NFGIEVAMIGYQCTVFPVLLAVWMSVKEKL 237  
QY 60 HKRLKGTADFLITPVLLLTGFTTFAIGPAMRWGVDVLAHGLQGLYDFGGPVGGLLFG 119  
DB 238 RRVIPDALDLITPFLVITSGFTALLIGPAGRALGDGTSFILSTLISHAGWLAGLFG 297  
QY 120 LVYSPVITGLHQSPFPIELFNQ---GGSFIFATASMANIAOGAACLAVFEFLAKSEKL 176  
DB 298 GLYSVIVITGIIHHSFAIEAGLLGNPSIGNVNFLLPIWAMANVAOGGACFAVWFKTDAKI 357  
QY 177 KGLAGSGVSAVLGITPAIFGVNLRWRPFFIGIGTAAIGGALLIALFNENIKAVAGLGAAGF 236  
DB 358 KAITLPSAFSAMLGITEAAIFGINLRFVKPFIALVGGAGGAWVSMHVMTAVGLTAI 417  
QY 237 LGVVSIDAPDMVPLCAVVTFFTAAGAAI 266  
DB 418 PGMAIVQASSLLNIIIGMAFAFAFAFALS 447

## RESULT 12

PTGA\_CORGL STANDARD; PRT; 674 AA.  
AC Q45298;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE PYS SYSTEM, GLUCOSE-SPECIFIC IIABC COMPONENT (EIIABC-GLC) (GLUCOSE-  
PERMEASE IIABC COMPONENT) (PHOSPHOTRANSFERASE ENZYME II, ABC  
COMPONENT) (EC 2.7.1.69) (EII-GLC/EIII-GLC).  
GN PTSG.  
OS Corynebacterium glutamicum (Brevibacterium flavum).  
OG Plasmid pBSG2.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Corynebacteriaceae;  
OC Corynebacterium.  
OX NCBI\_TaxID=1718;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 13869;  
RA Yoon K.-H.;  
RT "Cloning and nucleotide sequence of enzyme II of Brevibacterium  
lactofermentum phosphotransferase system.";  
RL Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.  
CC -I- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT  
SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE  
-TRANSPORT SYSTEM. THE IICD DOMAINS CONTAIN THE SUGAR BINDING SITE  
AND THE TRANSMEMBRANE CHANNEL; THE IIA DOMAIN CONTAINS THE PRIMARY  
PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HPR); IIA TRANSFERS ITS  
PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO  
THE SUGAR.  
CC -I- CATALYTIC ACTIVITY: PROTEIN N-PHOSPHOHISTIDINE + SUGAR =  
PROTEIN HISTIDINE + SUGAR PHOSPHATE.  
CC -I- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
CC -I- SIMILARITY: CONTAINS A PTS EIIA DOMAIN.  
CC -I- SIMILARITY: CONTAINS A PTS EIIB DOMAIN.  
CC -I- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.  
CC  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; L18875; AAA22992.1; -.  
DR HSP; P08837; 2F3G.  
DR InterPro; IPR001127; PTS\_EIIA.  
DR InterPro; IPR001996; PTS\_EIIB.  
DR InterPro; IPR003352; PTS\_EIIC.  
DR Pfam; PF00358; PTS\_EIIA\_1; 1.  
DR Pfam; PF02378; PTS\_EIIB; 1.  
DR Pfam; PF02378; PTS\_EIIC; 1.  
DR ProDom; PD001476; PTS\_EIIB; 1.  
DR ProDom; PD002243; PTS\_EIIA; 1.  
DR PROSITE; PS00371; PTS\_EIIA\_1; 1.  
DR PROSITE; PS1035; PTS\_EIIB\_CYS; 1.  
KW Phosphotransferase system; Sugar transport; Transferase;  
KW Phosphorylation; Transmembrane; Plasmid.  
FT DOMAIN 1 43 EIIA DOMAIN.  
FT DOMAIN ? ? EIIA DOMAIN.  
FT DOMAIN 542 674 EIIA DOMAIN.  
FT TRANSMEM 126 146 POTENTIAL.  
FT TRANSMEM 162 182 POTENTIAL.  
FT TRANSMEM 193 213 POTENTIAL.  
FT TRANSMEM 225 245 POTENTIAL.  
FT TRANSMEM 260 280 POTENTIAL.  
FT TRANSMEM 303 323 POTENTIAL.  
FT TRANSMEM 344 364 POTENTIAL.  
FT TRANSMEM 377 397 POTENTIAL.  
FT TRANSMEM 409 429 POTENTIAL.  
FT TRANSMEM 442 462 POTENTIAL.  
FT TRANSMEM 493 513 POTENTIAL.  
FT MOD\_RES 28 28 PHOSPHORYLATION (BY SIMILARITY).  
FT MOD\_RES 594 594 PHOSPHORYLATION (BY SIMILARITY).  
SQ SEQUENCE 674 AA; 71626 MW; 67A75AAF76E42FA2 CRC64;

Query Match 19.6%; Score 462.5; DB 1; Length 674;

Best Local Similarity 30.0%; Pred. No. 2.2e-23;

Matches 139; Conservative 79; Mismatches 204; Indels 41; Gaps 9;

QY 27 SLFGDLVAQAQYQCTVPLVSVSWILATIEKFLHKRLKGTADFLITPVLLLTGFTTFI 86  
DB 218 TVFGLPVLNDYSGQVFPPLIAAIGLYWVEKALKKIPEAVQMVVFPFFSLLIMIPATF 277  
QY 87 AIGPAMRWGVDVLAHGLQGLYDFGGPVGGLVGLVSPVITGLHQSPFPIELFNQ-G 145  
DB 278 LLGPFPGVGVNGISSLLEAVNFPFISIVIPLLYPLVPLGLHWPPLNMIQNLNTLG 337  
QY 146 GSIFATASMANIAOGAACLAVFEFLAKSEKLGLAGAS--GVSAVL--GITEPAIFGVNL 201  
DB 338 YDFIQGPMGAWNFACFGLVTGVFLIALKEKNRAMRQVSLGGLAGLGGISEPSLYGLL 397  
QY 202 RLWRPFGIGTAAIGGALLIALFNENIKAVAGLGVVSDIDAPDMVLMELVCAVVTFFIA 261  
DB 398 RFKTYFRLPLPGCLVGVGVIMGIFDIKAYF---VFTSLTIPAMPDWIGVTVGIAAFA-- 452  
QY 262 FGAAIAYGLYLVRRNGSID-----PDATAAPV-----PAGTTKA 295  
DB 453 FTSMLLVLFDFYDSRDAERDEAKAQAQAAEQTNTPAAPAAPAPAPAGAAAGGAAGAT-A 511  
QY 296 EACAPAEFNSDSTIIQAPLTGEALIALSSVSDAMFASKLGSVAIVTKQLVSPVSGKI 355  
DB 512 VATKPLRAAGQLVEITSPLLEGHAVPLSEVPDPIFAAGKLGFGIAETGTGTVAPADATV 571  
QY 356 VVAPPSGHAFATKAEKDGNSVDILMHIGFTVNLNCTHFNPLKKGDEKAGGLICEFD 415  
DB 572 ILVQKSGHVALRLE-----SGVELLIHIGLDTVOLGGEKGVHVERKQKQKAGDPLITFD 627  
QY 416 IDAIAAGYEVTTPIVVSNNYKKTGPVNTYGLGEIEAGANLLN 458  
DB 628 PEFIRSKNPLITPVVVSNNANKEGIEVIEAQAQADATTTVIKV 670

RESULT 13

SACX\_BACSU







```
FT DOMAIN 1 ?
FT DOMAIN 2 473
FT MOD_RES 29
FT TRANSMEM 111 131
FT TRANSMEM 159 179
FT TRANSMEM 189 209
FT TRANSMEM 226 246
FT TRANSMEM 259 279
FT TRANSMEM 301 321
FT TRANSMEM 341 361
FT TRANSMEM 371 391
FT TRANSMEM 399 423
FT TRANSMEM 441 461
FT CONFLICT 126 127
FT CONFLICT 184 194
FT CONFLICT 187 187
FT CONFLICT 307 307
FT CONFLICT 429 473
FT SEQUENCE 473 AA; 51080 MW; 7437F8822B624944 CRC64;

Query Match 15.5%; Score 367; DB 1; Length 473;
Best Local Similarity 30.1%; Pred. No. 2.9e-17;
Matches 80; Conservative 63; Mismatches 109; Indels 14; Gaps 5;

QY 1 MAMVPSLVNGYDVATMAAGEMP-MMSLFGLDVAQAGYQGTVPVLVYVSWILATIEKFL 59
Db :| | | | : :| | :| | :| | :| | :| | :| | :| |
195 VTLVSPQLMAY-----LLGQQLPEWDFGFMFSIAKVGYQAQVIPALLAGLAGLVETRL 249
QY 60 HKRLKGTADFLITPVLTLTLTGLTFIATGPAMRWVDVLA---HGLOGLYDFGCGPVGG 115
Db :| | | | :| | :| | :| | :| | :| | :| | :| |
250 KRIVPDYLYLVVVPVCSLILAVFLAHALGPGFMIGDGVAFVAVRHLMTGSE---APIGA 306
QY 116 LLFGLVYSPVITGLHQSPFPIELELF-NQGGSFIFATASMANIAQGAACLAFFFLAKSE.174
Db ||| :| | | | :| | :| | :| | :| | :| | :| | :| |
307 ALFGFLYAPLVITGVHOTTLAIDLQMIQSMGCTPWPLIALSNIAGGSAVIGIISSRKH 366
QY 175 KUKGLAGASGVSAVLGITEPAIFGVNLRURWFFFIGTAAIGGALIALFNKAKVALGAA 234
Db : : :| | | | | :| | :| | :| | :| | :| | :| | :| |
367 NEREISVPAAISAWLGVTEPAMYGINLKYRFPMLCAMIGSGLAGLICGLNGVMANGIGVG 426
QY 235 GFLGVVSDAPDPMVFLVCVVTFEI 260
Db | | | | :| | :| | :| | :| | :| | :| | :| |
427 GLPGILSIQPSYQWQV PALAMATAIII 452
```

Search completed: March 21, 2002, 16:28:49  
Job time: 329 sec

**THIS PAGE BLANK (USPTO)**



Qy	61	KRLKGTADFLLITPVLTLTGLFTTTAIGPMRWGDVLAHGLQGVLDFGGPVGGLLFG	120
Dd	1	: :::        :   :  :        :  :  :	
Dd	254	KYLNDADTFETFTLLSVITGTUTTVGPALRFVSNGLTDLGLVLYNTLGALGMLVFEG	313
Qy	121	VYSPIVITGLHQSFPPPIELELFNQ-----GGSFATASMANIAGGAACIAVFFLAKSE	174
Dd	314	FYSAIWVTGLHQSFPAETLMITNYQHSGIGGDFIPFVAACANMAQAAGATFAILEVTNI	373
Qy	175	KLKLAGASGSVAVLGITTEPAIEGVNLRWPFFIGTAGTAIGGALIALFNKAVALGAA	234
Dd	374	KTKALAAPAGVSAILGITTEPALFEIGNLKLYPFPIALGASAIGSLFMGLFHVLAVLSGA	433
Qy	235	GFLGWSIDAPDMVMFLVCVAWTFFFIAFGAAIAYGYLLYRRNSIDDPDATAVPVPGATTK	294
Dd	434	GLIGFISIKAGYNQLQMISIFISFLIAFVVVTSYGRMEAK--SITKE-----K	480
Qy	295	AEAEPAAEFSNDSTIIQ-----APLTGEAIALSSVSDAMFASKGLSGVAIVPTKGQ	346
Dd	481	NKONATTYOPEKVIIIDPVKSGELLAPINGFVIPLSDVDSPFSEIMKGKIAIKPKSGE	540
Qy	347	LVPVSGKIYVAFPSGHAFVRTKAEDSGSNVDILMHIGDFTVNLCNTHFNPLKKOGDEVK	406
Dd	541	LFSPADGEIIIAYETCHAYGIKT----NGGEVLLHGIDTYSMNGNGFIQNKKVGQVKV	596
Qy	407	AGELLCEFDIDATAKAGYEVTTPVIVSN---YKTKGPNVNTYGLGEIEAGANLL	456
Dd	597	AGDLLGSFKDEEKKSGLDDTVIIVITNSASYNEILPLSEN--VDIKVGEKIL	647
RESULT	2		
Q9L8G6			
ID	Q9L8G6	PRELIMINARY;	PRT; 627 AA.
AC	Q9L8G6;		
DT	01-OCT-2000 (TrEMBLrel. 15, Created)		
DT	01-OCT-2000 (TrEMBLrel. 15, Last sequence update)		
DT	01-JUN-2001 (TrEMBLrel. 17, Last annotation update)		
DE	SCRA.		
GN	SCRA.		
OS	Clostridium acetobutylicum.		
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;		
OX	Clostridium		
NCBI_TaxID=1488;			
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=ATCC 824;		
RA	MEDLINE=N0391269; PubMed=10937490;		
RX	Tanqney M., Mitchell W.J.;		
RT	"Analysis of a catabolic operon for sucrose transport and metabolism		
RL	in Clostridium acetobutylicum ATCC 824.";		
RL	J. Mol. Microbiol. Biotechnol. 2:71-80(2000).		
DR	EMBL; AF205034; AAF35839.1; -		
DR	InterPro; IPR001127; PTS_EIIA.		
DR	InterPro; IPR001996; PTS_EIIB.		
DR	InterPro; IPR003352; PTS_EIIC.		
DR	InterPro; IPR001254; Trypsin.		
DR	Pfam; PF00358; PTS_EIIA_1; 1.		
DR	Pfam; PF00367; PTS_EIIB; 1.		
DR	Pfam; PF02378; PTS_EIIC; 1.		
DR	ProDom; PD002243; PTS_EIIA; 1.		
DR	PROSITE; PS00371; PTS_EIIA_1; UNKNOWN_1.		
DR	PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.		
SQ	SEQUENCE 627 AA; 67596 MW; B835A80238FCA436 CRC64;		

Qy	63	LKGTADFLITPVLLLTTLTFTTAIGPAMRWGDVLAHGLQGLYDFGPGVGGLLFGLVY	122
Dd	242	VPEALDILLAPFLTMLITGPFAWVVGPGGRFVGDEISLGLQTLTYNTTFPSGVLFGLY	301
Qy	123	SPIVITGHOSFPPIELELFGNQG---SFIFATASMANIAQAACLAVFLEAKSEKLGL	179
Dd	302	SLIVITGIHSHFAIEAGLANPAIHKNFELLPITWSMANVAOGAALAVFKTRDKKKSI	361
Qy	180	AGAGSVSAVLGITETPAIFGVNRLRWRPFEGIGTTAAIGGALIALFNIKAVALCAAGFLVG	239
Dd	362	AAPASFCLGITETPAIFGVNLRXTRKFFIAGALGGALGGGYIVFTKVMTAVGVGTGIPGI	421
Qy	240	VSIDAPVMVFLVCVVYTFIFAEGAATAYGLYLVRNRNGSIDPATAAPVPAGTTKAEAA	299
Dd	422	AIVAQGSFLNYIIAMI---LAFGGAFIIMVLGIKEEITEEDLNKETYNKDIKVEEVES	477
Qy	300	PAEFSNDSTIIQAPLTGEATLSVSVDMPASFASKLGSGVAIVPTKGQLVSPVSGKIIVAF	359
Dd	478	-----VSPVNGKVLLKNVPDKTFAEGLIGDIGVDGEDVEGVSPIDGTGVHVHF	527
Qy	360	PSGHAFARPKAEDGSNVVDILMHIGFDTVNLGNTHFNPLKKQGDENVKAGELLECFDIDA	419
Dd	528	ETKHAIAMSK----NGVEMLIHGIDTVKMEGNGPKSFINDGEVKKGDKLIQFDLDLV	583
Qy	420	KAAGYEVTTPTVVSNYKKTGPVNT	443
Dd	584	KEKAVSPIVLTVITNHEDMGFVNS	607
RESULT	3		
ID	O9KG19	PRELIMINARY;	PRT; 630 AA.
AC	O9KG19;		
DT	01-OCT-2000 (TrEMBLrel. 15, Created)		
DT	01-OCT-2000 (TrEMBLrel. 15, Last sequence update)		
DT	01-JUN-2001 (TrEMBLrel. 17, Last annotation update)		
DE	PTS SYSTEM, BETA-D-GLUCOSIDE-SPECIFIC ENZYME II, ABC COMPONENT.		
GN	BH0296.		
OS	Bacillus halodurans.		
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;		
OC	Bacillus/Staphylococcus group; Bacillus.		
OX	NCBI_TaxID=86665;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-C-125 / JCM 9153;		
RA	MEDLINE=20512582; PubMed=11058132;		
RA	Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,		
RA	Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,		
RA	Horiuchi K.;		
RT	*Complete genome sequence of the alkaliphilic bacterium Bacillus		
RT	halodurans and genomic sequence comparison with Bacillus subtilis.*;		
RL	Nucleic Acids Res. 28:4317-4331(2000).		
DR	EMBL; AP001508; BAB04015.1; -		
DR	InterPro; IPR001127; PTS_EIIA.		
DR	InterPro; IPR001196; PTS_EIIB.		
DR	InterPro; IPR003352; PTS_EIIC.		
DR	Pfam; PF003358; PTS_EIIA_1; 1.		
DR	Pfam; PF00367; PTS_EIIB; 1.		
DR	Pfam; PF02378; PTS_EIIC; 1.		
DR	ProDom; PD002243; PTS_EIIA; 1.		
DR	PROSITE; PS00371; PTS_EIIA_1; 1.		
DR	PROSITE; PS01035; PTS_EIIB_CYS; 1.		
SW	Complete proteome.		
KQ	SEQUENCE 630 AA; 66978 MW; 6F0218011686ADD5 CRC64;		

QY 62 RLKGTADFLITPVLTLLLTGFLTFIAIGPAMRWGVDLAHLGLOQLYDFGPGVGLLFLGV 121  
 Db 237 KIHEAVKVFPTLILLVIVPVVTLIIILGPIGVVLGNIGIASVIOEITFSPVLGAIVAGI 296  
 QY 122 YSPIVITGHLQSPPIEL-ELFNQGGSFIFATASMANIAQGAACLAFFFLAKSEKLGKLA 180  
 Db 297 WQVLVIFGIHWGIIPIILNLSVRGDEVIKAVAPAVFSQAGALGVMLTKNKKLKALA 356  
 QY 181 GASGVSAVIGITPAIFGVNLRWPFPIGTAAGGALIALFNKKAVALGAAGFLGVV 240  
 Db 357 GSTSITALGITEPAVGVTLPLKKEPFINAVISAAGVGAIVGHYGSVAVAPAGLLTIP 416  
 QY 241 SI---DAPDMVFLVCAVVTFFTAAGAAATAYGLYLVRNGSIDPDATAAPVPACTTKAEA 297  
 Db 417 IFYPDGRGFVAEIVAIISFVLA-----AVLYIIVGFKDPVDDDTLSNESSENEVKR 471  
 QY 298 EAPAEPSNSTIIOAPLTGEAIALSSVSAMFASGLGSGVAIVPTKGOLVSPVSGKIVV 357  
 Db 472 EDDKKEPSASEEIKPLKGEVVPVTEQDHFVSSGAMGKGVAPKPEGRVLVAPINGTVTS 531  
 QY 358 APFSGHAFVTRKAEDGNSVDILMHGTDTVNLNGTHFNPLKQGDDEVKAGELLCFEDID 417  
 Db 532 LFTKHAIGITS---DNGTEIPFHVHIDTVQLKGEHFTSFIEQGDVAAGDVLLEFDVE 587  
 QY 418 AIKAAAGVETTPVWSNYAKTGVPNTYGLGEIEAGANLLNVAK 460  
 Db 588 RITAAGDVITPVLIITNAKQFNSVQTTDKREVTSDDLIIHVIK 630

## RESULT 4

Q9KF90 ID Q9KF90 PRELIMINARY; PRT; 636 AA.  
 AC Q9KF90;  
 DT 01-OCT-2000 (TremBLrel. 15, Created)  
 DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)  
 DT 01-JUN-2001 (TremBLrel. 17, Last annotation update)  
 DE PTS SYSTEM, BETA-GLUCOSIDE-SPECIFIC ENZYME II, ABC COMPONENT.  
 GN BGLP OR BH0595.  
 OS Bacillus halodurans.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 OC Bacillus/Staphylococcus group; Bacillus.  
 OX NCBI\_TaxID=86665;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C-125 / JCM 9153;  
 RX MEDLINE=20512582; PubMed=11058132;  
 RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,  
 RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,  
 RA Horikoshi K.;  
 RT "Complete genome sequence of the alkaliphilic bacterium Bacillus  
 RT halodurans and genomic sequence comparison with Bacillus subtilis.";  
 RL Nucleic Acids Res. 28:4317-4331(2000).  
 DR EMBL; AP001509; BAB04314.1; -;  
 DR InterPro; IPR001127; PTS\_EI1A.  
 DR InterPro; IPR001996; PTS\_EI1B.  
 DR InterPro; IPR003352; PTS\_EI1C.  
 DR Pfam; PF00358; PTS\_EI1A.1; 1.  
 DR Pfam; PF02378; PTS\_EI1B; 1.  
 DR Pfam; PF02378; PTS\_EI1C; 1.  
 DR ProDom; PD001476; PTS\_EI1B; 1.  
 DR ProDom; PD002243; PTS\_EI1A; 1.  
 DR PROSITE; PS00371; PTS\_EI1A.1; 1.  
 DR PROSITE; PS01035; PTS\_EI1B; 1.  
 DR Complete proteome.  
 SQ SEQUENCE 636 AA; 68437 MW; 434C0B12311716F2 CRC64;

Query Match 28.5%; Score 673.5; DB 2; Length 636;  
 Best Local Similarity 34.0%; Pred. No. 1.8e-30;  
 Matches 163; Conservative 78; Mismatches 193; Indels 45; Gaps 9;

QY 2 AMVFPSLVNGYDVAAATMAAGEMPWSLF-----GLDVAQACQGTVLPLVLV 47

Db 183 SLVYPTLV-----VLTEGE-PLYTLFTGCTIFESPWHITFLGIPVILMSYATSVIPIL 234  
 QY 48 VSWILATIEKFLKRLKGTADFLITLVLTLLTGLTFTFAIGPAMRWGVDLAHLGLOGLY 107  
 Db 235 AAYFASKVEARLKKIIPDVVKTFVLPFFELLIVVPLTFFIVIGPIATWAGOLLGOFTLWVY 294  
 QY 108 DFGGPGVGLLFLGVYSPVITGLHQSPPIEL-ELFNQGGSFIFATASMANIAQGAACLA 166  
 Db 295 NLSPIITAGALFGGFWGVFIFGLHMLIPIAINNVLVQGGSDPVLAMVFAASFAQIGAVAA 354  
 QY 167 VFPLAKSEKLGLAGASGVSAVLGITEPAIFGVNLRWPFPIGTAAGGALIALFNI 226  
 Db 355 WLKIKQKQVKTLSVPFISGIFGVTEPAIYGVTLPLKRPFIISCIANAAGGAILGLFRS 414  
 QY 227 KAVALAAGFLGVSI-----DAPDMVNF-LVCAVVTFFTAAGAAIAYGLYLVRNGSIDP 281  
 Db 415 QGVYIIGGLGIFGIPSLHPADGMAGFWGIVIAVVAVF-LGFILTYLFLGKSGNASDEQ 473  
 QY 282 DATAAPVACTTKAEAPAEFSNDSIIIOAPLTGEAIALSSVSAMFASGLGSGVAIV 341  
 Db 474 TETKAHTSTGTGEKEE-----ISSPFGNSVITLSEIKDEAFSSGALGEGIAIE 521  
 QY 342 PTRGQVLVSPVSGKIVVAFPSGHAFVTRKAEDGNSVDILMHIGDTVNLNGTHFNPLKQ 401  
 Db 522 PSEGLFSPVSGMVTALYPTPHALGIT---DRGAELLHIIGLDIVOLDGKFFTAHTIQ 577  
 QY 402 GDEVKAGELLCFEDIDAIKAAAGYEVTPPIVWSNYKTKGVPNTYGLGEIEAGANLLNVAK 460  
 Db 578 GAQVKGDLLEIDFKEIKAAAGYAVTTPVIVTNHKGQGLFDTDKQVNAVAGDRLLLETR 636  
 RESULT 5  
 Q9KLT8 ID Q9KLT8 PRELIMINARY; PRT; 479 AA.  
 AC Q9KLT8;  
 DT 01-OCT-2000 (TremBLrel. 15, Created)  
 DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)  
 DT 01-JUN-2001 (TremBLrel. 17, Last annotation update)  
 DE PTS SYSTEM, SUCROSE-SPECIFIC IIIC COMPONENT.  
 GN VCA0653.  
 OS Vibrio cholerae.  
 OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.  
 OX NCBI\_TaxID=666;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=EL TOR N16961 / SEROTYPE O1;  
 RX MEDLINE=20406833; PubMed=10952301;  
 RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,  
 RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Unayam L.A.,  
 RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,  
 RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,  
 RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,  
 RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,  
 RA Fraser C.M.;  
 RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio  
 RT cholerae.";  
 RL Nature 406:477-483(2000).  
 DR EMBL; AE004395; AAF96554.1; -;  
 DR TIGR; VCA0653; -;  
 DR InterPro; IPR001996; PTS\_EI1B.  
 DR InterPro; IPR003352; PTS\_EI1C.  
 DR Pfam; PF00367; PTS\_EI1B; 1.  
 DR Pfam; PF02378; PTS\_EI1C; 1.  
 DR PROSITE; PS01035; PTS\_EI1B; 1.  
 DR Complete proteome.  
 SQ SEQUENCE 479 AA; 50292 MW; DC0F129229957E3C CRC64;

Query Match 27.9%; Score 659; DB 2; Length 479;  
 Best Local Similarity 49.3%; Pred. No. 8.4e-30;  
 Matches 133; Conservative 46; Mismatches 87; Indels 4; Gaps 1;

[illegible]

RESULT	6
Q99RQ0	
ID	Q99R00 PRELIMINARY; PRT; 480 AA.
AC	Q99R00;
DT	01-JUN-2001 (TEMBLrel. 17, Created)
DT	01-JUN-2001 (TEMBLrel. 17, Last sequence update)
DE	01-JUN-2001 (TEMBLrel. 17, Last annotation update)
DE	P.T.S SYSTEM, SUCROSE-SPECIFIC IIBC COMPONENT.
GN	SCRA OR SA2167.
OS	Staphylococcus aureus subsp. aureus N315.
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;
OX	Bacillus/Staphylococcus group; Staphylococcus.
NCBI_TaxID=158879;	
RN	[1]
RP	SEQUENCE FROM N.A.
RA	Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA	Cui L., Oguchi A., Aoki K.I., Nagai Y., Lian J., Ito T., Kanamori M.,
RA	Matsumaru H., Maruyama A., Murakami H., Hosoyama A., Mizutani-Ui Y.,
RA	Takahashi N.K., Sawano T., Inoue R.I., Kaito C., Sekimizu K.,
RA	Hirakawa H., Kuhara S., Goto S., Yabuzaki J., Kanehisa M.,
RA	Iwashita A., Oshima K., Furuya K., Yoshino C., Shiba T., Hattori M.,
RA	Ogasawara N., Hayashi H., Hiramatzu K.;
RT	"Whole genome sequencing of methicillin-resistant Staphylococcus
RT	aureus."
RL	Lancet 357:1225-1240(2001).
PL	ENBL; AP003137; BAB43469.1; -.
DR	Complete proteome.
KW	SEQUENCE 480 AA; 51232 MW; D88607F6E0AF2E84 CRC64;
SQ	

Query Match	27.9%	Score	659;	DB 2;	Length	480;			
Best Local Similarity	48.6%;	Pred.	No. 8.4e-30;						
Matches	139;	Conservative	53;	Mismatches	82;	Indels	12;	Gaps	5;
QY	1	MAMVFPISLVNGCYDVAATMAAGE-MPMWSLFGDLDAQAQCYQGTVLPVLVVVSWILATIEKFL	59						
Db	194	MILVHPSLSMGAYDPKPAVEAGAKAIPYWDVGLGHINQYGQGVLPMLVAAYILASIEKGL	253						
QY	60	HKRLLGTAADFILTPVLTKLLTGFITFAIGPAMRWVGDLAHGQLGYDFCGPVGGILLFG	119						
Db	254	RKVIPTVDLNLLTPLLSIFITAFITFSPVGPITRQLGWLSDGLTWLYEFGGAIGGLIFG	313						
QY	120	LVSYPVITGLHQSFPIELF-----NQGSFTTFATSMANIAGGAACLAFFFLAK--SE	174						
Db	314	LLXAPIVTGMHHSFIARETTLIADATKTGSGFTFPIATMSNVAQGGAAIAAFTTIKQN	373						
QY	175	KLKGLAGASGVSVLGTETPAIFGVNLRLRPWFPGICTRAIGCALTALFNKAVALCAA	234						
Db	374	KLKGVAAGISALLGTEPAMFGVNLRKPPTFGAIVGSGIGSAYTAFKKVAIALGTA	433						
QY	235	GFLGVSWIDA--PDVMNFLCAVVTFTTFAFGAAIAYGLYLVRNRGS	278						

Db	434	GLPGFISINPVHAGWLHYFVGMOTISFII-----AITVTLILSKKAN	475
RESULT	7		
Q9KJ80			
ID	Q9KJ80	PRELIMINARY;	PRT; 644 AA.
AC	Q9KJ80;		
DT	01-OCT-2000	(TREMBLrel. 15, Created)	
DT	01-OCT-2000	(TREMBLrel. 15, Last sequence update)	
DT	01-JUN-2001	(TREMBLrel. 17, Last annotation update)	
DE	BETA-GLUCOSIDE-SPECIFIC EII PERMEASE.		
GN	BGLP.		
OS	Streptococcus mutans.		
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;		
OC	Streptococcus.		
OX	NCBI_TaxID=1309;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=NG8;		
RX	MEDLINE=20340959; PubMed=10878120;		
RA	Core C.K., Cvitkovitch D., Bielweis A.S., Honeyman A.L.;		
RT	"A novel beta-glucoside-specific PTS locus from Streptococcus mutans		
RL	that is not inhibited by glucose."		
RL	Microbiology 146:1555-1563(2000).		
DR	EMBL: AF206372; AAF89975.1;		
DR	InterPro; IPR001127; PTS_EIIA.		
DR	InterPro; IPR001996; PTS_EIIB.		
DR	InterPro; IPR003352; PTS_EIIC.		
DR	Pfam; PF00358; PTS_EIIA_1; 1.		
DR	Pfam; PF00367; PTS_EIIB; 1.		
DR	Pfam; PF02378; PTS_EIIC; 1.		
DR	ProDom; PD002243; PTS_EIIA; 1.		
DR	PROSITE; PS00371; PTS_EIIA_1; UNKNOWN_1.		
SQ	SEQUENCE 644 AA; 63282 MW; 723B/FFBDD2794EB CRC64;		

Query Match	27.9%	Score	658.5	DB	2	Length	644
Best Local Similarity	33.8%	Pred.	No. 1.3e-29				
Matches	160	Conservative	89	Mismatches	196	Indels	29
Gaps	12						
Qy	2	AMVFPVLVNGYDV---	AATWAAGBM---	PWMS-LFGLDV--	AOAGYOGTVLPVLVLSWIL	52	
Db		: : :	: : :	: : :	: : :	: : :	: : :
Db	185	ALVYPNTAASTVKHPLYTLFQGT	LVISPIYSFFGPIIFPASSYLQTVLPVVAIWAG	244			
Qy	53	ATIEKFLHKLKRGADPELLTPVLTLLT	GTFTTAIGAPMRWGDVLAHGLQGLYDRGP	112			
Db		: :	: :	: :	: :	: :	: :
Db	245	SKIETFEKKIIPDVVKVYVVPFFET	ILITVPLSELVFIQVPMWSADLVGAIFTGIYGNPV	304			
Qy	113	VGGLFLGLVTSPIVITGHQSGFDPPIE	-LELFNQGSGSFIFATSMANITAOCAACLA	171			
Db		: :	: :	: :	: :	: :	: :
Db	305	IYGVILGAMQVLYVMFLGHGLVPLAILE	LQKGGVILVATIAIC-FAQAGSLLINMRT	363			
Qy	172	KSEKLKGLAGASGVSAVLGITEPAIFGVN	ILRWPPFFIGTAAIGGALIALFNKRAVAL	231			
Db		: :	: :	: :	: :	: :	: :
Db	364	NENKROLSIPAFISALFGVTEPAIYGIT	ILPMRVPPFIMTCVSGAISGAYLALEFNVMQVM	423			
Qy	232	GAAGFLGVSDIAPD----	MVMFLVCAVVTFFETAFGRAAIYGL-YLVRNRS	286			
Db		: :	: :	: :	: :	: :	: :
Db	424	GGMLPAIPSFIDPKNSMILIHFIATAMNF	VLGVLTQFIKIPYLIGEFTSTDSVDK	483			
Qy	287	PVPAGTTIKAEAEPAEFSNDSTII	QAPLTGEATLSVSDAMFASGKLGSVAIVPTKG	346			
Db		: :	: :	: :	: :	: :	: :
Db	484	EEPVKELK-----	EIKQE--IISPLICKVVKLVNPDEVFASGAMGXIAIDPDGI	534			
Qy	347	LVPVSCKIVVAPSPGHAFVVRKAEDGS	NVDVILMHIGTDLNLTGHFNPLKKQGDVK	406			
Db		: :	: :	: :	: :	: :	: :
Db	535	VWAPTKGEVTLVPTKHAUGLRT--	ENGA--EILIHIGMDTVSLAGKFKFSFVEVGHDVE	590			
Qy	407	AGELLCCFFDIDAKAAGYEYTTPI	VWSNYRKKTKGPVNTYIGLGEIAGCANLINVAK	460			
Db		: :	: :	: :	: :	: :	: :
Db	591	AGOTLLSEFDYNAIKAAGLPVITPV	INSDOFEDVLTTOERTVEAGNYLLTAVK644				



```
RESULT 8
Q99X04 ID Q99X04 PRELIMINARY; PRT; 674 AA.
AC Q99X04;
DT 01-JUN-2001 (TremBLrel. 17, Created)
DT 01-JUN-2001 (TremBLrel. 17, Last sequence update)
DE PUTATIVE PTS SYSTEM ENZYME II.
GN SPY2097.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SF370;
RX MEDLINE=21192684; PubMed=11296296;
RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
RA Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,
RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,
RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
RT *Complete genome sequence of an M1 strain of Streptococcus pyogenes."
RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
DR EMBL; AE006630; AAK34748.1; -.
KW Complete proteome.
SQ SEQUENCE 674 AA; 72490 MW; 4BAA26B40E784878 CRC64;

Query Match 27.4%; Score 647.5; DB 2; Length 674;
Best Local Similarity 34.4%; Pred. No. 5.8e-29;
Matches 155; Conservative 86; Mismatches 178; Indels 31; Gaps 11;

QY 1 MAMVPSLVNGYDVAAATMAA--GEMPMWSLGLDVAQAQYQGTVPVLVSVWILATIEKFLH 58
DB 212 ICLVSPQLLNAYAVAGTAAETAKNMWVDFGFTINRIGYQAQVFPALLAGLSLAYLEIF 271
QY 59 LHKRGKTADFLLTPVLTLLTGLFTFIAIGPAMRWVGDVLAHGLQGLYDFG--GPVG-- 114
DB 272 WRKRIPVVMFMFVPLSLIPALIAHTVLGP----IGWTICKGISFVVLAGLTGPVKWL 327
QY 115 -GLLGLVYSPVITGLHOSFPPIELELFNOGG---SFIFATASMANIAQGAACLAFFVL 170
DB 328 FGAIFGALYAPLVTGLHMTNAIDTQLIADTATRTTGLWPMIALSNIAQGSAAVFAIYLM 387
QY 171 AKSEKLG-LAGASGVSALVIGTEPAIFGNLRLRPFFIGTGAIGGALIALFNKAV 229
DB 388 NRHEEREISUPAALSAYLGVTEPAIFGNVRYVFPVAGMISGIALGLSTTFNQAN 447
QY 230 ALGAAGFLGWSIDAPDMVFLVC----AVTFFIAFGAAIAYGLVLRNGSIDPDATA 285
DB 448 SIGVGLPGFMALNVKYMIPFFICMAVAIVPMFLTFPFRKSHIMTKTDEAKL-PETPV 506
QY 286 APVPAGTTKAEAPAEAFSNDSTI-IOAPLTGEAIALSSVDAMFASGLSGVAIVPTK 344
DB 507 SDAPVAT-----APHKTMQGTVTITLSTPLTGEVKALSEAVDPVPAQGVMGQALLQPT 560
QY 345 QGLVSPVSKIVVAPPSSGHAFVTRKAEDGSNVDIIMHIGFTDVTNLTGTHFNPLKQGB 404
DB 561 GVLVAPCDAEVSLFPTKHAICLVTT---EGLELHMIGMDVTNLDGGQFEALVRQGDQ 616
QY 405 VRAGELLCFDFDIDAIAKAGYEVTTPIVWSN 434
DB 617 VRAGQTILQIFDIAAISEAGYATETPLVVTN 646

RESULT 9
Q99Y91 ID Q99Y91 PRELIMINARY; PRT; 620 AA.
AC Q99Y91;
DT 01-JUN-2001 (TremBLrel. 17, Created)
DT 01-JUN-2001 (TremBLrel. 17, Last sequence update)
DE PUTATIVE SUCROSE-SPECIFIC PTS PERMEASE, ENZYME II.
GN SCRA OR SPV1815.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SF370;
RX MEDLINE=21192684; PubMed=11296296;
RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
RA Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,
RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,
RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
RT *Complete genome sequence of an M1 strain of Streptococcus pyogenes."
RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
DR EMBL; AE006608; AAK34540.1; -.
KW Complete proteome.
SQ SEQUENCE 620 AA; 65725 MW; 64F5FE83524DC4DC CRC64;

Query Match 26.7%; Score 632; DB 2; Length 620;
Best Local Similarity 35.6%; Pred. No. 3.9e-28;
Matches 160; Conservative 66; Mismatches 202; Indels 22; Gaps 6;

QY 1 MAMVPSLVNGYDVAAATMAAAGEMPMWSLGLDVAQAQYQGTVPVLVSVWILATIEKFLH 60
DB 189 LMLVSNELPNWVVA---SGGDKVPLTFPGF-VPVVGYQGTVPVPAFFVGLVGAKLEKWLH 244
QY 61 KRLKGTADFLLTPVLTLLTGLFTFIAIGPAMRWVGDVLAHGLQGLYDFGPPVGLLFLGL 120
DB 245 KKVPEALDLLVTPFTFAIMSTLGLFVIGPVPHSLNENLVLAGTQAVLHLPLFGIAGLIVGG 304
QY 121 VYSPVITGLHOSFPPIELELF-NOGGSFIFATASMANIAQGAACLAFFLAKSEKLG 179
DB 305 IQGLIVVTGIHIFNFLEAQLIANTGKDPFNAYLTAAATAAQAGATLAVAVTKTKLKL 364
QY 180 AGASGVSALVIGTEPAIFGNLRLRPFFIGTGAIGGALIALFNKAVKAAAGFLGV 239
DB 365 AEPSTLSALLGITEPAIFGNLRYPKVFSGLIGGALGGVWAGLFGIAGTGFITVLPGT 424
QY 240 VSIDAPDMVFLVCAYVTFEIAFGAAIAYGLVLRNGSIDPDATAAPVPAGTTKAEAEA 299
DB 425 LLYLNGQLLQVLTMLVGLGVAFAIATWGY-----QDRETLPLPAVEDQTDQ 474
QY 300 PAFSNDSTIIQAPLTGEAIALSSVDAMFASGLSGVAIVPTKQGLVSPVSGKIVVAF 359
DB 475 PALAEE---TLYSPLNGTVVDLSAVSDPVFSSGAMGGGLAIPKPEDNTLYSPVDGKVEIF 531
QY 360 PSGHAFVTRKAEDGSNVDIIMHIGFTDVTNLTGTHFNPLKQGBDEVKAGELLCFDFDIDA 419
DB 532 ETGHAIATIS-----SOGAEVLLHIGIDTSMAGDGFESLVAVGQAVKKGDLGHDFDSKI 587
QY 420 KAAGYEVTTPIVVSNNYKKTGPVNTYGLGEI 449
DB 588 AEAGLDTTMMIVSNIAQYQSDVILAAGHV 617

RESULT 10
Q48408 ID Q48408 PRELIMINARY; PRT; 621 AA.
AC Q48408;
DT 01-NOV-1996 (TremBLrel. 01, Created)
DT 01-NOV-1996 (TremBLrel. 01, Last sequence update)
DE CELLOBIOSE-SPECIFIC PTS PERMEASE.
GN KAS.
OS Klebsiella oxytoca.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Klebsiella.
OX NCBI_TaxID=571;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-P2;
```

```

RX MEDLINE=97176370; PubMed=9023916;
RA Lai X., Davis F.C., Hespell R.B., Ingram L.O.;
RT "Cloning of cellobiase phosphoenolpyruvate-dependent
RT phosphotransferase genes: functional expression in recombinant
RT Escherichia coli and identification of a putative binding region for
RT disaccharides.";
RL Appl. Environ. Microbiol. 63:355-363(1997).
DR EMBL; U61727; AAB51563.1; -
DR HSP; P20166; IGPR.
DR InterPro; IPR001127; PTS_EIIA.
DR InterPro; IPR001996; PTS_EIIB.
DR InterPro; IPR003352; PTS_EIIC.
DR Pfam; PF00358; PTS_EIIA_1; 1.
DR Pfam; PF00367; PTS_EIIB; 1.
DR Pfam; PF02378; PTS_EIIC; 1.
DR ProDom; PD001476; PTS_EIIB; 1.
DR ProDom; PD002243; PTS_EIIB; 1.
DR ProSITE; PS00371; PTS_EIIA_1; 1.
DR ProSITE; PS01035; PTS_EIIB_CYS; 1.
SQ SEQUENCE 621 AA; 65492 MW; DF1DE2A2A7B81339 CRC64;

Query Match      25.98; Score 612.5; DB 2; Length 621;
Best Local Similarity 32.4%; Pred. No. 4.9e-27;
Matches 154; Conservative 89; Mismatches 181; Indels 51; Gaps 11;

QY 2 AMVPSLVNGYDVAAATMAAGEMPWMSLFLDVAQAGYGTGVLPLVSVWILATIEKFLHK 61
DB 182 ALTHPLMQAED--ASWAPGAATEYFL-GIPVTFINSSVIPILASWSCWIEKRSNA 238
QY 62 RLKGTADFLTPVLTLLTGLTFAIGPAMRWGVGDVLAHGLQGLYDFGGVGLLFLV 121
DB 239 ILSSMKNFPTAICLAVVPLTELLIIPVATLWSQLLANGYQLIYQVAPWLAGAANGAL 298
QY 122 YSPIVITGLHOSFPPIEL-ELFNQGSFIFATASMANIAQAACLAFFFLAKSEKLG 180
DB 299 WQCVIFGLHGLLPLMINNLAVLGHDSMMPMLLPVAVMGQVGAALGFLTRDRARQVLA 358
QY 181 GASGSVAVLGITEPAIFGVNLRWRPFIFIGITAAIGGALIALFNKAKVALGAAGFLGV 240
DB 359 GSAYSAGIFGVTEPAIYGLNPLRPPIFCGVYPIGAWVG-FSDSHVSTYSFGFNII 417
QY 241 S-----IDAPDMVFLVCAVVTFFIAGAAIAYGLVLRNSIDPDATAAPYAG 291
DB 418 TLAQIMPEGIDAT-----VMGGAAGMEASLII-----ACVLTAVAG 454
QY 292 TTKAEE-----APAEFSDSTIIQAPLTGEATLSSVDAMPASCKLGSVAIVPTKG 345
DB 455 LPRSSAQAAVVVAPASV-ND---ILAPMTGSLVALDQVPDSTFASGLLGQGVAIIPSVG 510
QY 346 QLVSPVSGKIVVAPPSGHAFVTRKAEDGSNVDILMHIGFTVNLNGTHFNPLKKGDEV 405
DB 511 KVIAFSGEASVIFQTKHAGILLS----DSGIELLIHVGDITVKLDGAPTAHVKEGDKI 566
QY 406 KAGELCEFFDIDAKAAGEVTPPIVVSNNYKTKGPVNTYTLGTEAEANLNNAK 460
DB 567 KAGDLLEFDRQALDAGYLDATPIIISNSDDFTLDWVSASAVDAQOPLLSVSR 621

RESULT 11
Q9L461 PRELIMINARY; PRT; 577 AA.
AC Q9L461
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE PERMEASE.
GN BGLP.
OS Lactobacillus plantarum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
OC Lactobacillus.
OX NCBI_TaxID=1590;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=B21;
RX MEDLINE=20263773; PubMed=10802183;
RA Marasco R., Salatiello I., De Felice M., Sacco M.;
RT "A physical and functional analysis of the newly identified bglGPT
RT operon of Lactobacillus plantarum.";
RL FEMS Microbiol. Lett. 186:269-273(2000).
DR EMBL; AJ250202; CAB71150.1; -
DR InterPro; IPR001127; PTS_EIIA.
DR InterPro; IPR001996; PTS_EIIB.
DR InterPro; IPR003352; PTS_EIIC.
DR Pfam; PF00358; PTS_EIIA_1; 1.
DR Pfam; PF00367; PTS_EIIB; 1.
DR Pfam; PF02378; PTS_EIIC; 1.
DR ProDom; PD002243; PTS_EIIA_1; 1.
DR ProSITE; PS00371; PTS_EIIA_1; 1.
SQ SEQUENCE 577 AA; 60877 MW; E8C877050103B5FC CRC64;

Query Match      25.7%; Score 607; DB 2; Length 577;
Best Local Similarity 32.6%; Pred. No. 9.2e-27;
Matches 147; Conservative 81; Mismatches 133; Indels 90; Gaps 12;

QY 2 AMVPSLVNGYDVAAATMAAGEMPWMSLFLDVAQAGYGTGVLPLVSVWILATIEKFLHK 61
DB 188 ALLYPTLVTVMNSTTL-----HFFGIPVPTTSTVPIILLAVWVLSYLEPVLDK 239
QY 62 RLKGTADFLTPVLTLLTGLTFAIGPAMRWGVGDVLAHGLQGLYDFGGVPGV-GLLFLGL 120
DB 240 LFPAAIRNIETPLFSLIIMVPLTLLVDGPIGGLIGNGLASGVMAIYNF-MPIGAGVINGA 298
QY 121 VYSPIVITGLHOSFPPIEL-ELFNQGSFIFATASMANIAQAACLAFFFLAKSEKLG 179
DB 299 FWQVFVIFGVHVTVPVLMNNIAKMGYDPLLPILSAVLSQAGALAVFLKSRQKKKAL 358
QY 180 AGASGSVAVLGITEPAIFGVNLRWRPFIFIGITAAIGGALIALFNKAKVALGAAG---- 235
DB 359 AGSFVTALFGITEPTIYGVTLKLRPFYCAVVGALGGAIL-----GAAGTHAS 408
QY 236 -----FLG-----VVSIDAPDMVFLVCAVVTFFIAGAAIAYGLVLRNG 277
DB 409 SFTLPSSLAVPTFLGHGFMGEVIGL-----IVAFLLGAILTYFFGFA----- 450
QY 278 SIDPDATAAPVPAGTTKAEAEAPAEFS-----NDSTIIQAPLTGEATLSSVDAMPASG 332
DB 451 -----RQQADYSVVSQSDNTVM-APVEGTIPLTSVHDEFASE 490
QY 333 KLGSGVAIVPTKGOLVSPVSGKIVVAPPSGHAFVTRKAEDGSNVDILMHIGFTVNLNG 392
DB 491 AMGKGLAIVPNKGVTPAVDGTITAVYPTGHAIGITA-----NSGAELIHIGINTVQLNG 546
QY 393 THENPLAKKGDEVKAGELCEFFDIDAKAAG 423
DB 547 QYETVMYKQNOVKRGDILLTKFDVDKIKTAG 577

RESULT 12
Q46129 PRELIMINARY; PRT; 616 AA.
AC Q46129
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE PTS-DEPENDENT ENZYME II.
GN ABGF.
OS Clostridium longisporum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1523;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B6405;
RX MEDLINE=98151780; PubMed=9491080;

```

RA Brown G.D., Thomson J.A.;  
 RT "Isolation and characterisation of an aryl-beta-D-glucoside uptake and  
 RT utilisation system (abg) from the gram-positive ruminal Clostridium  
 RT species C. longisporum.",  
 RL Mol. Gen. Genet. 257:213-218(1998).  
 [2]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=B6405;  
 RA Brown G.D., Thomson J.A.;  
 RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; L49336; AAC05713.1; -.  
 DR HSSP; P20166; IGPR.  
 DR InterPro; IPR001127; PTS\_EIIA.  
 DR InterPro; IPR001996; PTS\_EIIB.  
 DR InterPro; IPR003352; PTS\_EIIC.  
 DR Pfam; PF00358; PTS\_EIIA\_1; 1.  
 DR Pfam; PF00367; PTS\_EIIB; 1.  
 DR Pfam; PF02378; PTS\_EIIC; 1.  
 DR ProDom; PD001476; PTS\_EIIB; 1.  
 DR ProDom; PD002243; PTS\_EIIA; 1.  
 DR PROSITE; PS00371; PTS\_EIIA.1; UNKNOWN\_1.  
 DR PROSITE; PS01035; PTS\_EIIB\_CYS; 1.  
 SQ SEQUENCE 616 AA; 65890 MW; 7D9CD1B17BE9283E CRC64;

Query Match 24.9%; Score 589.5; DB 2; Length 616;  
 Best Local Similarity 31.2%; Pred. No. 9.8e-26;  
 Matches 146; Conservative 80; Mismatches 197; Indels 45; Gaps 9;  
 QY 2 AMVFPVLNGYDVAAATMAAGEMPMSLFLGDA--OAGYGGTGLVPLVSVWILATIEKFL 59  
 DB 181 AMVFP-----AITSFAGQIV--DFGIPVMPSSSQSVIPILAIYIATSKVEKLF 230  
 QY 60 HRLKKTADFLTPVLTLLTGLFTFIATGPMRWVGDVLAHGLQGLYDFGPGVGLLFG 119  
 DB 231 KKIIPDMVKTLFVPFATLLVVPVTFMPTIAANALGDLTLAIYNFNPTIAGLFIG 290  
 QY 120 LVYSPITVITGLHQSFPPIEL-ELFNQGGSFIFATASMANIAAGAACLAFFFLAKSEKLG 178  
 DB 291 GFQWQVFMFGLHGLVPIAMNNAVIGYDPVLATSVAVCFQGVVMAIAKTCKRKLK 350  
 QY 179 LAGASGVAVLGITEPAIFGVNLRWPFIFIGTAAIGGALIALFNIAKAVAGAGFLG 238  
 DB 351 LCIPALISGFGVTEPAIYGITLPRKKPILSCIAAGVTGGIIGFESKGYSGMLGIFA 410  
 QY 239 VVSIDAPD-----MMVFLCAVVTFFIAFGAATAYGLYLVRNGSIDDPATAAPVPA 290  
 DB 411 LPSYINPEGIDRGFYGMVIAVAGIVVGFIL-----MFVTKLNDSEVKTTSKKEE 462  
 QY 291 GTTKAEAPAEFSNDSTIIQAPLTGEATALSVDAMPASGKLGSGVAIVPTKGOLVSP 350  
 DB 463 SLVKQEE-----IVSPIQGEVVTIAEVKDEAFSSGALGKGVAINPEGKVYAP 510  
 QY 351 VSGKIVVAPPSGHAFVTRKAEDGSNVDILMHIGFTVNLNGTHFNPLKKQGDVEVKAGEL 410  
 DB 511 ADGTLTFLPSSLHALGITT--ENGA--EILIHVMDTVLEQKHFTAKVKQGDKIRKQGL 566  
 QY 411 LCEFDIDAIAKAGYEVTTPIVSNYKKTGPVNTYGLGETEAGANLNV 458  
 DB 567 LIEFOKEAIEKAGYSTITPVLITNSDQYLDVTETDKRKVDVNSLETV 614

RESULT 13  
 Q9CFK9 PRELIMINARY; PRT; 636 AA.  
 AC Q9CFK9;  
 DT 01-JUN-2001 (TRENBLrel. 17, Created)  
 DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)  
 DE BETA-GLUCOSIDE-SPECIFIC PTS SYSTEM IIBAC COMPONENT (EC 2.7.1.69).  
 GN PTBA.  
 OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;

OC Lactococcus.  
 OX NCBI\_TaxID=1360;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=IL1403;  
 RA Bolotin A., Wincker P., Mauger S., Jaillon O., Malarne K.,  
 RA Weissenbach J., Ehrlich S.D., Sorokin A.;  
 RT "The complete genome sequence of the lactic acid bacterium Lactococcus  
 RT lactis.",  
 RL Genome Res. 0:0-0(2001).  
 DR EMBL; AE006376; AAK05558.1; -.  
 DR InterPro; IPR003439; ABC\_transportr.  
 DR InterPro; IPR001127; PTS\_EIIB.  
 DR InterPro; IPR001996; PTS\_EIIB.  
 DR InterPro; IPR003352; PTS\_EIIC.  
 DR Pfam; PF00358; PTS\_EIIA\_1; 1.  
 DR Pfam; PF00367; PTS\_EIIB; 1.  
 DR Pfam; PF02378; PTS\_EIIC; 1.  
 DR ProDom; PD001476; PTS\_EIIB; 1.  
 DR ProDom; PD002243; PTS\_EIIA; 1.  
 DR PROSITE; PS00211; ABC\_TRANSPORTER; UNKNOWN\_1.  
 KW Complete proteome.  
 SQ SEQUENCE 636 AA; 67630 MW; 012030D819163325 CRC64;

Query Match 24.9%; Score 588.5; DB 2; Length 636;  
 Best Local Similarity 33.5%; Pred. No. 1.2e-25;  
 Matches 148; Conservative 76; Mismatches 185; Indels 33; Gaps 10;  
 QY 2 AMVFPVLNGYDVAAATMAAGEMPMSLFLG--DVAQAGYGGTGLVPLVSVWILATIEKFL 59  
 DB 182 ALVYPCFVDG--TVAKTFA--ESGGLSLFGIPFSPVPTAGYSSVMPITAITFAAFLEHQL 238  
 QY 60 HRLKKTADFLTPVLTLLTGLFTFIATGPMRWVGDVLAHGLQGLYDFGPGVGLLFG 119  
 DB 239 KKIIPDMVKTLFVPFATLLVVPVTFMPTIAANALGDLTLAIYNFNPTIAGLFIG 298  
 QY 120 LVYSPITVITGLHQSFPPIEL-ELFNQGGSFIFATASMANIAAGAACLAFFFLAKSEKLG 178  
 DB 299 GFQWQVFMFGLHGLVPIAMNNAVIGYDPVLATSVAVCFQGVVMAIAKTCKRKLK 358  
 QY 179 LAGASGVAVLGITEPAIFGVNLRWPFIFIGTAAIGGALIALFNIAKAVAGAGFLG 238  
 DB 359 LAIPAISGFGVTEPAIYGITLPRKKPILSCIAAGVTGGIIGFESKGYSGMLGIFA 415  
 QY 239 VVSIDAPD-----MMVFLCAVVTFFIAFGAATAYGL--YLVRRNGSIDDPATAAPVPA 295  
 DB 416 IFSFTSNITLTGVSQGAIKMIVSAVAVIAGVVTVLV-----GFEDDVNIENIP----- 465  
 QY 296 EAPAPAEFSNDST--IIQAPLTGEATALSVDAMPASGKLGSGVAIVPTKGOLVSP 352  
 DB 466 ----DKKFNKQTKNKEIGSLGKGVIPUSQVKDAAFSAGVMGKGAIEPTLGEVRAPFD 521  
 QY 353 GKTIVVAPPSGHAFVTRKAEDGSNVDILMHIGFTVNLNGTHFNPLKKQGDVEVKAGEL 412  
 DB 522 GVMVILFPFKHAGLLIS---NEGTELLIHIGIDTVQLEKGYFETFKVQKQSVKKGDL 577  
 QY 413 EFDIDAIAKAGYEVTTPIVSNV 434  
 DB 578 KFDIERIQNAGYSTQVPIVTN 599

RESULT 14  
 Q9KAS1 PRELIMINARY; PRT; 470 AA.  
 AC Q9KAS1;  
 DT 01-OCT-2000 (TRENBLrel. 15, Created)  
 DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)  
 DE PTS SYSTEM, TREHALOSE-SPECIFIC ENZYME II, BC COMPONENT.  
 GN BH2216.  
 OS Bacillus halodurans.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 22, 2002, 07:02:16 ; Search time 66.97 Seconds  
(without alignments)  
5163.982 Million cell updates/sec

Title: US-09-604-231-1  
Perfect score: 1527  
Sequence: 1 ctcattgccttcgcgcgtt.....gttgaaaccttgagtgttgcg 1527

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents\_NA: \*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq: \*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq: \*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq: \*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq: \*  
5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq: \*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	109.6	7.2	3615	1	US-08-920-812-17
C 2	109.6	7.2	3615	1	US-08-920-827-17
C 3	109.6	7.2	3615	1	US-08-921-177-17
C 4	109.6	7.2	3615	1	US-08-362-577C-17
C 5	109.6	7.2	3615	2	US-08-920-828-17
C 6	97.8	6.4	465	2	US-08-673-190A-3
C 7	71.6	4.7	357	2	US-08-673-190A-6
C 8	43.2	2.8	7218	1	US-08-232-463-14
C 9	37.2	2.4	4403765	4	US-09-103-840A-2
C 10	34.6	2.3	46899	1	US-08-471-119A-1
C 11	32.8	2.1	12912	2	US-08-460-751-1
C 12	32.8	2.1	14060	3	US-08-658-136-4
C 13	32.6	2.1	477	4	US-09-135-994-1
C 14	32.6	2.1	80161	3	US-09-036-987A-1
C 15	32.6	2.1	80161	4	US-09-370-700-1
C 16	32.4	2.1	292	4	US-09-117-121-29
C 17	32.2	2.1	7218	1	US-08-232-463-14
C 18	32	2.1	1008	4	US-09-198-955A-9
C 19	32	2.1	1901	1	US-08-153-848-43
C 20	32	2.1	1901	3	US-09-299-843A-43
C 21	32	2.1	1901	5	PCT-US93-11153-43
C 22	32	2.1	2232	1	US-08-241-766-12
C 23	32	2.1	2453	5	PCT-US95-07180-1
C 24	32	2.1	3120	1	US-08-491-146-1
C 25	32	2.1	3120	1	US-08-241-766-11
C 26	32	2.1	3120	1	US-08-234-011-1
C 27	32	2.1	3120	2	US-08-701-062A-1

28 32 2.1 3533 4 US-08-985-916-15 Sequence 15, Appl  
C 29 32 2.1 5163 3 US-08-700-651-1 Sequence 1, Appl  
C 30 32 2.1 5163 3 US-08-928-361B-4 Sequence 4, Appl  
C 31 32 2.1 5318 3 US-08-700-651-2 Sequence 2, Appl  
C 32 32 2.1 5318 3 US-08-928-361B-3 Sequence 3, Appl  
C 33 32 2.1 10348 2 US-08-457-273B-41 Sequence 41, Appl  
C 34 32 2.1 10348 3 US-08-556-419-13 Sequence 13, Appl  
C 35 32 2.1 10348 4 US-09-041-886-14 Sequence 14, Appl  
C 36 32 2.1 10366 1 US-08-246-982A-5 Sequence 5, Appl  
C 37 32 2.1 10366 1 US-08-453-265-5 Sequence 3, Appl  
C 38 31.8 2.1 1227 3 US-09-074-912-3 Sequence 2, Appl  
C 39 31.6 2.1 5511 3 US-08-928-361B-2 Sequence 1, Appl  
C 40 31.6 2.1 7334 3 US-08-928-361B-1 Sequence 1, Appl  
C 41 31.6 2.1 7655 1 US-08-619-554-1 Sequence 196, App  
C 42 31.2 2.0 342 4 US-08-905-223-196 Sequence 429, App  
C 43 31.2 2.0 818 4 US-08-998-416-429 Sequence 4, Appl  
C 44 31 2.0 34303 2 US-08-735-609-4 Sequence 4, Appl  
C 45 31 2.0 34303 2 US-08-735-609-4 Sequence 4, Appl

#### ALIGNMENTS

RESULT 1  
US-08-920-812-17/c  
; Sequence 17, Application US/08920812  
; Patent No. 5763188  
; GENERAL INFORMATION:  
; APPLICANT: Ohno, Tsuneya  
; APPLICANT: Matsuhisa, Akio  
; APPLICANT: Uehara, Hirotugu  
; APPLICANT: Eda, Soji  
; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease  
; NUMBER OF SEQUENCES: 25  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 6300 Sears Tower, 233 South Wacker Drive  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States of America  
; ZIP: 60606-6402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC Compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION NUMBER: US/08/920,812  
; FILING DATE: 29-AUG-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/362,577  
; FILING DATE: 27-MAR-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Rin-Laures, Li-Hsien  
; REGISTRATION NUMBER: 33,547  
; REFERENCE/DOCKET NUMBER: 19036/32420  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312/474-6300  
; TELEFAX: 312/474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 17:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3615 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: Genomic DNA  
; ORIGINAL SOURCE:  
; ORGANISM: Escherichia coli  
; STRAIN: Clinical Isolate EC-24  
US-08-920-812-17

Query Match	7.28;	Score	109.6;	DB 1;	Length	3615;			
Best Local Similarity	46.3%;	Pred.	No. 2.2e-23;						
Matches	540;	Conservative	0;	Mismatches	589;	Indels	37;	Gaps	4;
Qy	312	tcactccagtgctga	cgctgctc	accggatctc	ttaacattcat	cgccagttgcccag	371		
Db	2495	TCACACCAATGCTAT	GTCTGATGGTAT	TCACACCGTCA	CCCTTCTGCTGTGGGGCGGC	2436			
Qy	372	caatgcgctgggtg	ggcgatgctg	ggcacacggtct	acagggactttat	gatcttcggtg	431		
Db	2435	TATCAACCTGTGA	AGCAAGTATG	TGCGCGGTAT	CTCTGGCTTTAT	CAGGGGTC	2376		
Qy	432	gtccagtgccgggt	ctgctcttc	ggcttggctt	ggctactcaacaat	gctgaactggtctgc	491		
Db	2375	CTGATTTGCGGG	CGCGGTAA	TGGCGGCTTCT	TGGCAATCTTCG	TCATGTTCCGAGTC	2316		
Qy	492	accagtcctccgc	ccaattgagct	ggagctgtt	taacacgggtggat	cccttcattctcg	551		
Db	2315	ACTGGGCTTGGT	GCCGGTGTAT	CAATAACTT	TCACCGTCTGGGCTAC	GACACCATGA	2256		
Qy	552	caacggcatctatg	---gcta	atcgcacagg	tgggcgatgtttgg	cagtgcttctcc	608		
Db	2255	TCCGCGCTTAA	TGCCCGCCAT	TATGCGCAG	GTGCGGCGCGCTG	CGCTCTCTCTCT	2196		
Qy	609	tggcgaaagtg	aaaaactca	aggcccttc	gagctgcttc	aggtgtctccgctattctg	668		
Db	2195	CGAAGCGGAT	GCACGA	AAAAAGTGG	TGGCGGATCAG	CGCGGTTGACGAGTCTGTGTG	2136		
Qy	669	gtattcggagctgc	gatcttc	gggttga	accttcgctgc	gtgcgtgcgttcttcacgc	728		
Db	2135	GTATACCGA	ACCAGCGGTAT	ATGGGTCA	ACCTGCCGCGTAAGT	TACCCCTTTGTTATCG	2076		
Qy	729	gtatcggtacgc	agcatcgtg	ggcgcttg	gattgcaactctt	aaatacaagcagttg	788		
Db	2075	CCTGTATCAG	TGGGCGCTT	TGGGGCCAC	CAATATTGGGTAC	GCGCAACGAAAGTCTACT	2016		
Qy	789	cgtttggcgctgc	aggtttctt	gggtgtgtt	ctctattgat	gtccagatggtgtcagt	848		
Db	2015	CCTTTGGTTT	GCCAAGTATTT	TCACTTCAT	GTAAACCATCC	CGCTCAACGGGAATTGATT	1956		
Qy	849	tcttgggtgtg	caagttgt	tacctcttc	atcgcatcggc	agcagattgcttatggcc	908		
Db	1955	TCACCGTCTG	GGCGAGCGT	TATGCGCGTG	TCATTGGCATCGGTT	TCGCGATTTTCGGTA	1896		
Qy	909	tttacttggttc	gcgcgaac	ggcagcttg	atcagatgca	accgctgctccagtgccctg	968		
Db	1895	CGGTGATGCT	TCAATTTAT	CACCGCTA	ACGTACCGCAG	CGCAGGGTGCCCC-----	1842		
Qy	969	caggaaacgac	caaaagc	gaagcaga	aacccgcagaa	atttcaaaagattccaccatca	1028		
Db	1843	-----GA	AGAAAAAC	CACAGAGT	TATTACAC	CACTTACGAGCGGGCGGTA	1798		
Qy	1029	tccaggcac	ctttga	ccggtgaag	ctattgca	ctgagcagctcagcagatcccatgtttg	1088		
Db	1797	TCTGTTT	CACCATAC	CGGGAGAGAT	TGTGCGCTCAT	TTCAGTCTGCTGATAC	1738		
Qy	1089	ccagcgaaa	agcttggctc	ggcggttgc	atcgtccca	accaagggcgagttattctc	1148		
Db	1737	CCAGTGG	CGCTTTTGGG	TAAAGGTAT	TGCCATTTCT	CGCCTCGGTTGGTGAAGTGC	1678		
Qy	1149	cggtagtg	gaaagatgt	ggtggcatt	cccatctg	gccatgcttgcagcttcgcacca	1208		
Db	1677	CGGTTG	CGGGTCGA	AAATTTGCT	TGTTTGGCC	ACATATACA-----GCCCAATTGGCATG	1624		
Qy	1209	aggcttg	aggatgg	ttocaa	tgttgatctt	gtatgcacattggtttcgacacagtaacc	1268		
Db	1623	AGTCAG	TATG	-----TGT	GGAGTCTGTAT	TCATGCTGCGTATCGACACCGTAAAC	1570		
Qy	1269	tcaacgc	acgcac	ctttaa	ccccgctga	agcaggcgatgaagtcaacagcagggagc	1328		
Db	1569	TGGACG	CAAAATCTTTT	TCCGCTC	CACGTC	CAAGTGGGTGCAAGCTCAATACAGCGCATC	1510		

```

Qy 1329 tgctgtgtgaattcgattatgatgcattattaaagctgcagggttatgaggtaaccacgcgcga 1388
|||||  ||| |||||  ||| |||  ||| |||  ||| |||  ||| |||||
Db 1509 GCGTGAATTTCTTTGTATATCCCTGCTATTTCGCGAGCGCGGATTTGATCTGACGACGCCGG 1450
|||||  ||| |||||  ||| |||  ||| |||  ||| |||  ||| |||||
Qy 1389 ttgttgttctgaattacaagaaaaacgcgcacctgtgtaaacacttaccggtttggcgaaattg 1448
|||||  ||| |||||  ||| |||  ||| |||  ||| |||  ||| |||||
Db 1449 TATTAAATCACTAATACGGATGATTTTACGAGCAGTATTACCCACGCGACGCCGAGATAA 1390
|||||  ||| |||||  ||| |||  ||| |||  ||| |||  ||| |||||
Qy 1449 aagcgggagccaaactgctcaacgct 1474
|||||  ||| |||||  ||| |||  ||| |||  ||| |||  ||| |||||
Db 1389 GCGCAGGTGAACCGCTGTTATCCATC 1364
|||||  ||| |||||  ||| |||  ||| |||  ||| |||  ||| |||||

RESULT 2
US-08-920-827-17/c
: Sequence 17, Application US/08920827
: Patent No. 5770375
: GENERAL INFORMATION:
: APPLICANT: Ohno, Tsuneya
: APPLICANT: Matsuhisa, Akio
: APPLICANT: Uehara, Hirotosugu
: APPLICANT: Eda, Soji
: TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
: NUMBER OF SEQUENCES: 25
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
: STREET: 6300 Sears Tower, 233 South Wacker Drive
: CITY: Chicago
: STATE: Illinois
: COUNTRY: United States of America
: ZIP: 60606-6402
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/920,827
: FILING DATE: 29-AUG-1997
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/362,577
: FILING DATE: 27-MAR-1995
: ATTORNEY/AGENT INFORMATION:
: NAME: Rin-Laures, Li-Hsien
: REGISTRATION NUMBER: 33,547
: REFERENCE/DOCKET NUMBER: 19036/32420
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 312/474-6300
: TELEFAX: 312/474-0448
: TELEX: 25-3856
: INFORMATION FOR SEQ ID NO: 17:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3615 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: Genomic DNA
: ORIGINAL SOURCE:
: ORGANISM: Escherichia coli
: STRAIN: Clinical Isolate EC-24
: US-08-920-827-17

```

	Query Match	7.2s;	Score 109.6;	DB 1;	Length 3615;
	Best Local Similarity	46.3s;	Pred. No. 2.2e-23;		
	Matches 540;	Conservative	0;	Mismatches 589;	Indels 37; Gaps 4;
Qy 312	tcaactccagtgctgacgtgtgctgtcaccggatctcttacattcatcgccattggccag	371			
Db 2495	TCACACCATTTCTATGCTGTATGGTTATFACACACCGTCACCTTTCTGCTGGTGGGCGC	2436			
Qy 372	caatgcactggatggcgaatgtgtgcacacacgctctcacacggacatttatgatcttcgtg	431			

Db 2435 TATCAACCTGGGATAGCGAACTGATTGCGCGGTTATCTCTGGCTTTATCAGCGGGTTC 2376  
QY 432 gtccagtcggcggtctgctcttcggtctggtctactcaaccaatgcatacactgctgc 491  
Db 2375 CTGCATTTTCGGGCGCGGTAATGGCGGCTTCTGCGAAATCTTCGTCAATGTTGCGACTGC 2316  
QY 492 accagtcctcccgcaaatgagctggagctgtttaaccagaggtgagctcctcatctctcg 551  
Db 2315 ATGGGGCTGTGGCGGTGTGTATCAATAACTTCACTGCTGGCTGAGACACCATGA 2256  
QY 552 caacggcatctatg---gctaataatgccaggggtgctgcatgtttggcagtggtttcc 608  
Db 2255 TCCCGCTGTAATGCCCGCATATATGGCGCAGTTCGGGCGCGCTCGCGCTTCTCCTCT 2196  
QY 609 tggcgaagagtgaaagctcaaggccctgcaaggctgttcaggtgtctccgctgttcttg 568  
Db 2195 GCGAAGCGGATGCGCAAGAAAGTGGTGGCGGGATACGCGCGGTTCAGCAGTCTGTTTG 2136  
QY 669 gtattacggagctgcgctcttcggtgtgaaccttcgctgcgctgcccgtttcttcacg 728  
Db 2135 GTATCACCGAACCAGCGGTATATGGGTCACTGCGCGGTAAAGTACCCCTTTGTTATCG 2076  
QY 729 gtatcggtaccgcagctatcggtggcgcttattgacactctttaataatcaaggcagttg 788  
Db 2075 CCTGTATCATCTGGGCTTTGGGGCGCACCATATTGGCTACGCGCAACGAAAGTCTACT 2016  
QY 789 cgttggcgctgcaggtttcttgggtgtgtttcttctattgatctcagatatggtcatgt 848  
Db 2015 CCTTTGGTTTGCAGTATTTTACCTTCACTGCAACCACTCCCGTCAACGGGAATTGATT 1956  
QY 849 tcttgggtgtcaggtgttactcttctcatcgcttcgctggcgagcaggtgttctatggcc 908  
Db 1955 TCACCGTCTGGCCACCGTTATTTGGCGGTGTCATTCGCCATCGGTCGGCATTTGCGGTA 1896  
QY 909 tttacttggctgcgcgcaacgcagcagcattgatccagatgcaacccgctgcctcagtgctg 968  
Db 1895 CGGTGATGCTTTCATTTTCATCCCGTAAACGTCAGCGCAGCGAGGTGCCCC----- 1842  
QY 969 caggaaacgcaaaagccgaagcagagcaccgcgagaaatttcaaacgattccaccatca 1028  
Db 1843 -----GAAGAGAAACACCAAGAGTATTATACACCCTGAGCAGGCGGTA 1798  
QY 1029 tccagggcactttagcgttgaagctattgcaactgaagcagcagcagcagcagcagcagc 1088  
Db 1797 TCTGTTACCGATGACGGGAGAGATGTCGCTCATTCACGTCGCTGATACACAGCTTTG 1738  
QY 1089 ccagcggaagcgttgctggcggttgccatcgctcccaacgaaggcgagtgagttcttc 1148  
Db 1737 CCAGTGGCCTGTTGGTAAAGGTATTGCCATTCTGCCCTCGGTGGTGAAGTCCGTTCTC 1678  
QY 1149 cggtagtggaagattgtgtggtgattcccatctgcccattgttgcagttcgcaacca 1208  
Db 1677 CGGTTGCGGGTTCGAATGCTGCTGTTGTCGCGCACATTTACA-----CGCCATTGGCATTTG 1624  
QY 1209 aggtcagagattgttccaatgttgatacttgaatcacaattggtttcagacagtaaac 1268  
Db 1623 AGTCAGATATGG-----TGTGGAGATCCTGATTTCATGTCGGTATFCGACACCGTAAAC 1570  
QY 1269 tcaacggcagcagcactttaaccgctgaagaagcagggcgatgaagtcaagcagggagc 1328  
Db 1569 TGGACGGCAAAATCTTTTCGGCTCAGCTCAGCTCAAGTGGGTGACAAAGTCAATACAGGGATC 1510  
QY 1329 tgcgtgtgaatgcagatatgtagcattgaagctgaggtgttatgaggttaacacagcgga 1388  
Db 1509 GGCTGATTTCTTTTGTATATCCCTGCTATTTCGGGAGCGCGGATTTGATCTGACGACGCCGG 1450  
QY 1389 ttggtgttcgaattacaagaaacggacctgttaaacacttacggttttggcgcaaatgg 1448  
Db 1449 TATTAATAGTAAATACGATGATTTTACGGACGTATTATACCCACGCGCACGGCGAGATAA 1390  
QY 1449 aagcgagagccaacctgtccaacgtc 1474

Db 1389 GCGCAGGTGAACCGCTGTTTATCCATC 1364  
RESULT 3  
US-08-921-177-17/c  
; Sequence 17, Application US/08921177  
; Patent No. 5798211  
; GENERAL INFORMATION:  
; APPLICANT: Ohno, Tsuneya  
; APPLICANT: Matsunisa, Akio  
; APPLICANT: Uehara, Hirotugu  
; APPLICANT: Eda, Soji  
; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease  
; NUMBER OF SEQUENCES: 25  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 6300 Sears Tower, 233 South Wacker Drive  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States of America  
; ZIP: 60606-6402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/921,177  
; FILING DATE: 29-AUG-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/362,577  
; FILING DATE: 27-MAR-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Rin-Laures, Li-Hsien  
; REGISTRATION NUMBER: 33,547  
; REFERENCE/DOCKET NUMBER: 19036/32420  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312/474-6300  
; TELEFAX: 312/474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 17:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3615 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: Genomic DNA  
; ORIGINAL SOURCE:  
; ORGANISM: Escherichia coli  
; STRAIN: Clinical Isolate EC-24  
; US-08-921-177-17  
  
Query Match 7.2%; Score 109.6; DB 1; Length 3615;  
Best Local Similarity 46.3%; Pred. No. 2.2e-23;  
Matches 540; Conservative 0; Mismatches 589; Indels 37; Gaps 4;  
  
QY 312 tcaactcagtcgctgagctgtgctgcacggattccttaccattcgcgcatcgccag 371  
Db 2495 TCACACCAATTGCTATGCTGATGGTTATCACACCGCTCACCTTCTGCTGGTGGGCGCGC 2436  
QY 372 caatgcgctgggtggcgatgtgctggcacaggtgtacagggactttatgattcggctg 431  
Db 2435 TATCAACCTGATAAGCGAACTGATTGCGCGGTTATCTCTGGCTTTATCAGCGGGTTC 2376  
QY 432 gtccagtcggcggtctgctcttcggtctgtactcaccacactgcatacactggtctgc 491  
Db 2375 CTGCATTTTCGGGCGCGGTAATGGGCGGCTTCTGGCAAAATCTTCGTCAATGTTGCGACTGC 2316  
QY 492 accagtcctcccgcaaatgagctggagctgtttaaccagaggtgagctgtttaaccaggggtgagcttcttcg 551  
Db 2315 ACTGGGGCTGTGGCGGTGTGTATCAATAACTTCACTGCTGGCTGAGACACCATGA 2256





```
QY 729 gtatcggtaccgcagctatcggtggcgctttgattgcactctttaataatcaaggcagttg 788
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2075 CCTGTATCAGTGGGCTTTGGGGCCACCACATATTGGCTACGCCGAACGAAGTCTACT 2016
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 789 cgttggcgctgcaggtttcttgggtgttcttattgatgtccagatatggtcatgt 848
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2015 CCTTTGGTTTGGCAAGTATTTCACCTTCATGCAAAACCATCCCGTCAACGGGAATTGATT 1956
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 849 tcttggtgtcagttgttacctcttcctcatcgcttcgcttcgagcgagcttattgagcc 908
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1955 TCACCGCTCTGGGCCAGCGTATTGGGGGTGTCATTGCCATCGGTTGCGCATTTGTCGGTA 1896
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 909 ttacttggtttcgcgcgaacgcagcattgatccagatcaaacccgtctgcagtgccgtg 968
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1895 CGGTGATGCTTCATTTTCATCACCGGTAACGTCAGCCACGCGAGGTGCCCC----- 1842
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 969 caggaaacacaaagccgaagcagacgacccgcgagaattttcaaacgattccaccatca 1028
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1843 -----GAAGAGAAACACAGAGGTTATTACACCACCTGAGCAGGGCGGTA 1798
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1029 tccaggcacctttgacgggtgaagctattgactgagcagctcagcagtgccatgtttg 1088
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1797 TCTGTTACCGGATGACGGGAGAGATTGTCTGCTCATTCACGCTCGCTGATACACGTTTG 1738
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1089 ccagcggaagccttgctcgggcttgccatcgctcccaaccaaggcgagttagtttctc 1148
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1737 CCAGTGGCCCTGTGGTAAAGGTATTGCCATTCTGCCCTCGGTTGGTGAAGTCGGTTCTC 1678
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1149 cggtagtggaagatttgggtggcattcccatctggccatgctttcgcagttcgcacca 1208
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1677 CGGTTGCGGGTCGAATTGCTTCTGTTCCGCCACATTTACA-----CGCCATTGGCATTG 1624
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1209 aggtcagagatggttccaattggttatatttgatgcacattggtttcgacagtaaac 1268
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1623 AGTCAGATGATGG-----TGTGGAGATCCTGATTCTCGGTATCGACACCGGTAAAC 1570
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1269 tcaacgcgcagcacttaaccgcgtgaagaagcagcgcatgaagtaaaagcagggagc 1328
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1569 TGGACGGCAAAATCTTTTCGGCTCAGCTCAGCTGGGTGACAAAGTCAATACAGGCGATC 1510
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1329 tgcgtgtgaattcgatattgattgccattaaaggctgcaggttatgaggttaaccacgcga 1388
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1509 GGCTGATTTCTTTTGATATCCCTGCTATTTCGGCAGGCGGATTTGATCTGACGACGCGG 1450
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1389 ttgtgtttcgaaattacaagaaacgcgacctgttaaacacttacggtttggcgcaaatg 1448
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1449 TATTAATCAGTAATAGCATGATTTTACGGACGATATTACCCACGCGACGCGCAGATAA 1390
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1449 aagcgaggagccaaacctgctcaacgtc 1474
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1389 GGCAGGTGAACCGTGTTATCCATC 1364
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

## RESULT 5

```
US-08-920-828-17/c
; Sequence 17, Application US/08920828
; Patent No. 5853998
; GENERAL INFORMATION:
; APPLICANT: Ohno, Tsuneya
; APPLICANT: Matsuhisa, Akio
; APPLICANT: Uehara, Hirotugu
; APPLICANT: Eda, Soji
; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESS: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```

```
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/920,828
; FILING DATE: 29-AUG-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/362,577
; FILING DATE: 27-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Rin-Laures, Li-Hsien
; REGISTRATION NUMBER: 33,547
; REFERENCE/DOCKET NUMBER: 19036/32420
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3615 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Escherichia coli
; STRAIN: Clinical Isolate EC-24
; US-08-920-828-17
```

Query Match 7.2%; Score 109.6; DB 2; Length 3615;

Best Local Similarity 46.3%; Pred. No. 2.2e-23; Mismatches 589; Indels 37; Gaps 4;

Matches 540; Conservative 0;

```
QY 312 tcaactcagtgctgacgttgctcaccggattctctacattcatcgccattggccag 371
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2495 TCACACCAATTGCTATGCTGATGGTTATCACACCGCTCACCTTTCTGCTGGTGGCGCG 2436
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 372 caatgcgctggggtggcgatgtgctggcgacacgggtctacaggacatttatgattcggtg 431
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2435 TATCAACCTGGATAAGCAACTGATTGCCGCGGTTATCTCTGGCTTTATCAGCGGTTTC 2376
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 432 gtccagtcggcggtctgctctcggtctgttactacccaatcgatcactcactcggtctgc 491
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2375 CTGCATTTGGGCGCGCGGTAATGGCGGCTTCGGCAAAATCTTCGTCATCTTCGGAATGC 2316
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 492 accagtccttccgcgaattgagctggagctgtttaaccagggtgagctcttcatcttcg 551
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2315 ACTGGGGCCTGGTGGCGGTGTGTATCAATAACTTCACCGTGTCTGGGCTACGACACCATGA 2256
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 552 caacggcatctatg---gctaatacgcccgaggtggtggcgatgtttggcgagtttcttc 608
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2255 TCCCGCTGTTAATGCCCGCATTTATGGCGCAGGTCGCGGCGCGCTCGCGGCTCTCTCT 2196
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 609 tggcgaagagtgaaagctcaaggccttgcaggtgttcaggtgttcctcgctgttccttg 668
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2195 GCGAACGCGATGCGCAGAAAAAGTGGTGGCGGATCAGCGGCGTTCACAGAGTCTGTTTG 2136
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 669 gtattacggagcctgcgactcttcggtgtgaaccttcgctgcgctggcggttcttcacg 728
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2135 GTATCACCAGAACCGGATATATGGCGTCAACCTGCCGCGTAAGTACCCCTTTGTTATCG 2076
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 729 gtatcggtaccgcagctatcggtggcgctttgattgatgcactcttataatacaagcgagttg 788
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2075 CCTGTATCAGTGGGCTTTGGGGCCACCACATTTATGGCTACGCCGAACGAAGTCTACT 2016
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 789 cgttggcgctgcaggtttcttgggtgttcttattgatgtccagatatggtcatgt 848
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2015 CCTTTGGTTTGGCAAGTATTTCACCTTCATGCAAAACCATTCGCCGTCAACGGGAATTGATT 1956
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 849 tcttggtgtgcagttgttacctcttcctcatcgcttcgagcgaggtattgcttatggcc 908
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Db	1955	TCACCGCTGGGCCACCGTTATTGGGGGTGTCATTGGCATCGGTTGCCATTGTCGGTA	1896
Qy	909	tttacttggttcgcgcaacgcgcagcattgatccagatgcacacgctctccagtgcctg	968
Db	1895	CGGTGATGCTTCATTTTCATCACCGCTAAACGTCACCCAGCGCAGGGTCCCCC-----	1842
Qy	969	caggaacgcaccaaagccgaagcagaagcaccgcgcagaattttcaaacgattccaccatca	1028
Db	1843	-----GAAGAGAAACACCAGAGGTTATTACACCACCTGAGCAGGGCGGTA	1798
Qy	1029	tccaggcacctttgacgcgtgaagctattgcactgacgagcagcgtcagcagatccatgtttg	1088
Db	1797	TCGTGTCACCGATGACGGAGAGATTGTGCTCCTATTTCACGTCGCTGATACCACGTTTG	1738
Qy	1089	ccagcggaagcttggctcggcgttgccatctgcctcccaacaaagggcgagttattctc	1148
Db	1737	CCAGTGGCCTGTGGGTAAAGTATTGGCATTTCTGCCCTCGTTGGTGAAGTGCCTCTC	1678
Qy	1149	cgtgagtggaagattggttgccattcccatctcctcctcctgcccagtttcgcagctcgcacca	1208
Db	1677	CGGTTGCGGGTCGAATTGCTGCTGTTTGGCCACATTACA-----CGCCATTGGCATTG	1624
Qy	1209	agctgagatggtttccaatgttggaatatcttgatgcacatggttttcgacacagtaaac	1368
Db	1623	AGTCAGATGATG-----TGTTGAGATCCGTGATTCTATGTCGGTATCGACACCGCTAAAC	1570
Qy	1269	tcaacggcacgcacttaaccgcgtgaagacgagggcgatgaagtcaaacgagggagc	1328
Db	1569	TGACGGCCAAATTTCTTTCCGCTCAGTCAACGTGGGTGACAAGTCAATACAGCGATC	1510
Qy	1329	tgctggtgaattcgatattgatccattaagctgcaggttatgaggttaaccacgcgcga	1388
Db	1509	GGCTGATTTCTTTTGATATCCCTGCTATTCGCGAGGCGGATTTGATCTGACGACGCGG	1450
Qy	1389	tgtgttttcgaattacaagaataaacgcctgtaaaacacttcagttttgggcgaataat	1448
Db	1449	TATTATCAGTAATACCGATGATTTTACGGACGTATTACCCACGGCACGGCCAGATAA	1390
Qy	1449	aagcgggagccaaactgctcaacgtc	1474
Db	1389	GCACGGTGAACCGGTGTTATCCATC	1364

```

RESULT      6
US-08-673-190A-3/c
; Sequence 3, Application US/08673190A
; Patent No. 5985668
;
; GENERAL INFORMATION:
; APPLICANT: Mattes, Ralf
; APPLICANT: Klein, Kathrin
; APPLICANT: Stegmaier, Sabine
;
; TITLE OF INVENTION: Sucrose Metabolism Mutants.
;
; NUMBER OF SEQUENCES: 14
;
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner
;
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA

```

TELECOMMUNICATION INFORMATION:		
TELEPHONE:	(202)408-4000	
TELEFAX:	(202)408-4400	
INFORMATION FOR SEO ID NO:	3:	
SEQUENCE CHARACTERISTICS:		
LENGTH:	465 base pairs	
TYPE:	nucleic acid	
STRANDEDNESS:	both	
TOPOLOGY:	linear	
MOLECULE TYPE:	DNA (genomic)	
ORIGINAL SOURCE:		
ORGANISM:	Protaminobacter rubrum	
US-08-673-190A-3		

Query Match	6.4%;	Score 97.8;	DB 2;	Length 465;
Best Local Similarity	53.5%;	Pred. No. 2.7e-20;		
Matches 204;	Conservative 0;	Mismatches 177;	Indels 0;	Gaps

Qy	145	cgccaccatggctggcgggcgaaatgccaatgtgtgcctctgttggttttagatgttgc	204
Db	441	CGCTGGGGGCTCGGGCGGTTTCCACACATGAATTTCTTCGGACTCGATATCGCCAT	382
Qy	205	agccggttaccaggacccgtgcttctctgtgctgggtgttcttggattcttggcaacgat	264
Db	381	GATCGGTATCAGGGTACGGTGTTCGCGTGCTGCTACGGTGTGGTGTATGACGGTGCT	324
Qy	265	cgagaagtctctgcacagcgactcaaggcgactgcagacttctctgatactccagtgct	324
Db	321	GGAAACACAGTCGGCAGAGTGATCCCAACGGCGCTGGACCTGATCCTGACGCCATTCCT	262
Qy	325	gacgttctgtcaccggattcccttaattcatgcgcattggccccagcaatgcgtgggt	384
Db	261	GACGGTGGTGATCTCCGGCTTCGTGCGCATGCTGTTATCGGCCCGCGCGGCGCACACT	202
Qy	385	ggcggatgtctggcacacaggtctacagggaacitattgatatttcgggtgttcagtcggcg	444
Db	201	GGGTGACGGCATCTCTTCCTGTGTGACACGCTGATTGCCACGCGCGGTGGTTCGCCGG	142
Qy	445	tctgctcttgggtgtggtctactacacgaatcgctcaactcgttctgcaccagtccttccc	504
Db	141	ATTCTCTTCGGCGGCTGTATTCCGGCCATCGTTCATCACCGGCATTCACACACGCTTCCA	82
Qy	505	gccaattgagctggagctgtt	525
Db	81	CGCGGTAGAGGGGGGCTGCT	61

RESULT 7  
US-08-673-190A-6  
; Sequence 6, Application US/08673190A  
; Patent No. 5985668  
; GENERAL INFORMATION:  
; APPLICANT: Mattes, Ralf  
; APPLICANT: Klein, Kathrin  
; APPLICANT: Stegmaier, Sabine  
; TITLE OF INVENTION: Sucrose Metabolism Mutants.  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: Finnegan, Henderson, Farabow, Garrett & Dunner  
; STREET: 1300 I Street, N.W., Suite 700  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20005-3315  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patencin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/673,190A  
; FILING DATE: 27-JUN-1996

```

; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Forman, David S.
; REGISTRATION NUMBER: 33,694
; REFERENCE/DOCKET NUMBER: 06473.0001-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)408-4000
; TELEFAX: (202)408-4400
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 357 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Enterobacter species
US-08-673-190A-6

Query Match          4.7%; Score 71.6; DB 2; Length 357;
Best Local Similarity 53.1%; Pred. No. 2.4e-12;
Matches 152; Conservative 0; Mismatches 134; Indels 0; Gaps 0;

QY 183 ttttgggttagattgtgccaaagcgggttacaggccacgtgtctctgtggtgag 242
Db 67 TCTTGGCATCAAGTGGCGATGATCGGTACACGGCACCGCTCTTCCCGGTGCTGCTGG 126

QY 243 ttcttggatttggcaacgatcgagaagtctctgcacaagcgactcaaggcgactgcag 302
Db 127 CGGTGTGTTTATGACATGTCGAAACGCTGCGSSCGTTATCCCTGACGCGCTGG 186

QY 303 atttctgactactcagtgctgaagtgctgctcaccggattcttaccatcagcca 362
Db 187 ACTGATCTCACCCCGTTCCTGACGGTGATATCTCGGCTTTATCGGCCCTGCTGCTGA 246

QY 363 ttggcccaacaatgctggtggcgagtgctggcacacggtctcacaggacattatg 422
Db 247 TCGGCCCGCCGCTCGCGGCTCGCGAGCGTATTCGTTTATCTCATCAGCAGCTTATCA 306

QY 423 atttgggtggtccagtcggcggtctgctcttcgtggtggtctactc 468
Db 307 GCCACGCCGCTGGCTGGCGGCGCTGCTGTTGGGGCGCTCTACTC 352

RESULT 8
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232.463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935.313

```

```

; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-Fls
US-08-232-463-14

Query Match          2.8%; Score 43.2; DB 1; Length 7218;
Best Local Similarity 7.1%; Pred. No. 0.0072;
Matches 30; Conservative 207; Mismatches 185; Indels 0; Gaps 0;

QY 147 ccaccatgctgcggcggaatgccaaatgctccctgctggtggttctggaattctggaacgacg 206
Db 1034 CGAGCTTGCTGCAGTCGAGGAGCTTCGATATATATATATATATATATATATATATATATAT 1093

QY 207 ccggttaccaggcgaccgtgctctctgctggtggtggtggtggtggtggtggtggtggtggt 266
Db 1094 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1153

QY 267 agaagtctctgcacaagcgactcaaggcgactgacttctgactgactgactgactgactgactg 326
Db 1154 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1213

QY 327 cgttgcgtcaccggattcctacattcattcattcattcattcattcattcattcattcattcatt 386
Db 1214 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1273

QY 387 gcgatgctgcgcacacgctctacagggacttatgattcgtggtcagtcgagtcgagtcgagtc 446
Db 1274 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1333

QY 447 tgctcttcgtgctgctactaccacaaatcgctacatcgctgctgctgctgctgctgctgctgct 506
Db 1334 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1393

QY 507 caattgagctggagctgtttaaccagggtggtcattcattcattcattcattcattcattcatt 566
Db 1394 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1453

QY 567 ct 568
Db 1454 CT 1455

RESULT 9
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24

```

```
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match          2.4%; Score 37.2; DB 4; Length 4403765;
Best Local Similarity 43.7%; Pred. No. 18;
Matches 165; Conservative 0; Mismatches 213; Indels 0; Gaps 0;

QY 627 tcaaggccctgcaggtgctccaggtgtctccgctgttcttggtattacagagccctgcga 686
Db 3923279 TTACCGCGCTGCGCGCGCTTGCGCGCGTCCCGCGATGCCCTGGCTGCCAGCGTTACCG 3923320

QY 687 tcttcggtggaaccttcgctgcgtgcgttcttccttccttccttccttccttccttccttc 746
Db 3923219 CCGGTAGTGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 39233160

QY 747 tcttcggtggaaccttcgctgcgtgcgttcttccttccttccttccttccttccttccttc 806
Db 3923159 CCGTTCGCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 39233100

QY 807 tcttcggtggaaccttcgctgcgtgcgttcttccttccttccttccttccttccttccttc 866
Db 3923099 CCGACGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 39233040

QY 867 ttaccttcttccttccttccttccttccttccttccttccttccttccttccttccttcct 926
Db 3923039 CCACCGTTCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3922980

QY 927 acggcagcattgatccagatgcaacgctgctccagtgctgcaggaacaccaaagccg 986
Db 3922979 GTCCCGCGCGTTCGCCCGCGTTCGCCCGCGTTCGCCCGCGCGCGCGCGCGCGCGCG 3922920

QY 987 aagcagaagcaccgcgag 1004
Db 3922919 AAGTCGCGGAGCGCGCG 3922902

RESULT 10
US-08-471-119A-1/c
; Sequence 1, Application US/08471119A
; Patent No. 5827706
; GENERAL INFORMATION:
; APPLICANT: Leitner, Ernst
; APPLICANT: Schneider, Elisabeth
; APPLICANT: Schoergerdorfer, Kurt
; APPLICANT: Weber, Gerhard
; TITLE OF INVENTION: Cyclosporin Synthetase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5827706artis Corporation
; STREET: 59 Route 10
; CITY: East Hanover
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07936
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/471,119A
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
```

```
; ATTORNEY/AGENT INFORMATION:
; NAME: Kassenoff, Melvyn
; REGISTRATION NUMBER: 26,389
; REFERENCE/DOCKET NUMBER: 100-8029/CONT/CONT
; TELEPHONE: 201 503 8474
; TELEFAX: 201 503 8807
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 46899 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Tolypocladium niveum
; STRAIN: ATCC 34921
; US-08-471-119A-1

Query Match          2.3%; Score 34.6; DB 1; Length 46899;
Best Local Similarity 55.4%; Pred. No. 9.4;
Matches 67; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY 61 cgttcggcgccgaatgagttccttggtggcgccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 120
Db 29376 CGATACGGCGGCTCTGAGCTGCTGGAGTGGCGCGTTCGTCAGGGTTCGAGTTCGCGAC 29317

QY 121 ctttggtgaacggtacgacgtggtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 180
Db 29316 CTTGATGGTTCGAGGAGTGTACAGGACAGCAGTGCCTCGCAATTTGCCATCGGTG 29257

QY 181 c 181
Db 29256 C 29256

RESULT 11
US-08-460-751-1/c
; Sequence 1, Application US/08460751
; Patent No. 5891628
; GENERAL INFORMATION:
; APPLICANT: Readers, Stephen
; APPLICANT: Schneider, Michael
; APPLICANT: Glucksmann, Sandra
; TITLE OF INVENTION: IDENTIFICATION OF POLYCYSTIC KIDNEY
; TITLE OF INVENTION: DISEASE GENE, DIAGNOSTICS AND TREATMENT
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/460,751
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/413,580
; FILING DATE: 03-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7638-005
```





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 22, 2002, 06:40:16 ; Search time 1607.29 Seconds  
(without alignments)  
10208.988 Million cell updates/sec

Title: US-09-604-231-1  
Perfect score: 1527  
Sequence: 1 ctcacggcatctgcgcgtt.....gttgaacacctgagtgttcg 1527

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues  
Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
EST:  
1: em\_estfun:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estcom:\*  
5: em\_estpl:\*  
6: em\_estba:\*  
7: em\_estro:\*  
8: em\_estov:\*  
9: em\_htc:\*  
10: gb\_estl1:\*  
11: gb\_estl2:\*  
12: gb\_htc:\*  
13: gb\_gss:\*  
14: em\_gss\_fun:\*  
15: em\_gss\_hum:\*  
16: em\_gss\_inv:\*  
17: em\_gss\_pin:\*  
18: em\_gss\_pro:\*  
19: em\_gss\_rod:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query %	Match Length	ID	Description
c 1	48.8	3.2	680	10	AL506262
c 2	45.8	3.0	473	11	T03017
c 3	41.8	2.7	681	13	CNS02EOD
4	41.6	2.7	895	13	AL193990 Tetraodon
5	41.4	2.7	477	11	AL066286 Drosophila
6	41.4	2.7	559	10	BE997958 EST429681
7	41.4	2.7	601	11	AW574083 EST316674
8	41.4	2.7	627	11	BE980528 EST482254
9	40.8	2.7	884	13	BE998413 EST430136
c 10	40.8	2.7	983	13	AL065923 Drosophila
11	40.8	2.7	1101	13	CNS03CVD
c 12	40.2	2.6	970	13	AL078875 Drosophila
					AL243904 Tetraodon

13	39.8	2.6	606	11	BE997957
c 14	39.2	2.6	993	13	CNS00LOP
c 15	38.8	2.5	539	10	AI070146
c 16	38.8	2.5	714	10	AW583970
c 17	38.6	2.5	401	10	AI436816
c 18	38.6	2.5	1204	11	BE997958
c 19	38.4	2.5	1101	13	CNS0021D
c 20	38.2	2.5	500	11	BE997957
c 21	38.2	2.5	525	13	CNS025EN
c 22	38.2	2.5	531	10	AW573719
c 23	38.2	2.5	621	11	BE998387
c 24	38.2	2.5	652	10	AI982977
c 25	38.2	2.5	687	10	AW980715
c 26	38.2	2.5	703	10	AW980735
c 27	38.2	2.5	713	11	BE981619
c 28	38.2	2.5	737	11	BE981892
c 29	38.2	2.5	789	11	BE982423
c 30	38.2	2.5	795	11	BE983630
c 31	37.8	2.5	654	10	AI812147
c 32	37.6	2.5	780	11	BE980735
c 33	37.6	2.5	1101	13	CNS00FKE
c 34	37.4	2.4	460	10	AI318021
c 35	37.4	2.4	552	10	BE721200
c 36	37.4	2.4	904	13	CNS03HAO
c 37	37.2	2.4	421	10	AW173566
c 38	37.2	2.4	561	10	AI917973
c 39	37.2	2.4	594	10	AW440291
c 40	37.2	2.4	622	10	AI499080
c 41	37.2	2.4	647	10	BE158004
c 42	37.2	2.4	650	10	AW516253
c 43	37.2	2.4	710	10	AI564728
c 44	37.2	2.4	906	13	CNS02GJN
c 45	37	2.4	440	13	AZ982613

#### ALIGNMENTS

RESULT 1  
AL506262/c  
LOCUS AL506262 680 bp mRNA EST 04-JAN-2001  
DEFINITION AL506262 Hordeum vulgare Barke developing caryopsis (3.-15.DAP)  
Hordeum vulgare cDNA clone HY02124T 5', mRNA sequence.  
ACCESSION AL506262  
VERSION AL506262.1 GI:12032477  
KEYWORDS EST.  
SOURCE barley.  
ORGANISM Hordeum vulgare  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae  
; Triticeae; Hordeum.  
REFERENCE 1 (bases 1 to 680)  
AUTHORS Michalek W., Weschke W., Pleissner K.-P. and Graner A.  
TITLE EST sequencing and analysis in barley  
JOURNAL Unpublished (2000)  
COMMENT Contact: Michalek W  
Institute for Plant Genetics and Crop Plant Research  
Corrensstr.3, D-06466 Gatersleben, Germany  
Email: michalek@ipk-gatersleben.de, http://pgrc.ipk-gatersleben.de  
Seq primer: T3 primer for 5'end.

#### FEATURES

Location/Qualifiers  
1..680  
/organism="Hordeum vulgare"  
/cultivar="Barke"  
/db\_xref="taxon:4513"  
/clone="HY02124T"  
/clone\_lib="Hordeum vulgare Barke developing caryopsis (3.-15.DAP)"  
/tissue\_type="developing caryopsis (3.-15.DAP)"  
/lab\_host="XLOR"  
/note="Vector: plasmid pBK-CMV; Site.1: EcoRI; Site.2: XhoI; mRNA was made from developing caryopsis (3.-15.DAP) of spring barley variety 'Barke', a high quality malting





REFERENCE 3 (bases 1 to 681)  
 AUTHORS Genoscope.  
 TITLE Direct Submission  
 JOURNAL Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases  
 COMMENT This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at <http://www.genoscope.cns.fr/Tetraodon>.

FEATURES  
 source Location/Qualifiers  
 1. .881  
 /organism="Tetraodon nigroviridis"  
 /db\_xref="taxon:99883"  
 /clone\_lib="262H14"  
 /clone\_lib="G"  
 /note="Genoscope sequence ID : C0A262DD07LP1-end : T7"

BASE COUNT 74 a 105 c 170 g 282 t 50 others  
 ORIGIN

Query Match 2.7%; Score 41.8; DB 13; Length 681;  
 Best Local Similarity 40.9%; Pred. No. 2.2;  
 Matches 139; Conservative 18; Mismatches 183; Indels 0; Gaps 0;

QY 632 ggcctgcagggtcctcaggtctccgctgtcttctggtattacggagcctgcgactcttc 691  
 Db 58 GCGGTGTGTGTTTGTGTTGTTGTCGCGYGTGTTGTTGTTTGTGTTGTT 117  
 QY 692 ggtgtgaacctgcctgcgcgtgcccgtttcttcattcattcggtatccgcagcattcgt 751  
 Db 118 GTTGTGTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 177  
 QY 752 ggcgccttgatgcactctttaaatacaaggcagttgcgttggcgctcaggtttcttg 811  
 Db 178 GTNGTGTGTTGTCGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 237  
 QY 812 ggtgtgtttctattgatctccagatatggtcattgtcttctggtgtgcagttgttacc 871  
 Db 238 GTCCGCGTGTGTTGTCYGTGTTTGTGTCGTCGTCGTCGTCGTCGTCGTCGTC 297  
 QY 872 ttcttcacgcattcggcgcagcagttgattgaccttacttgccttcgcgcgaacgac 931  
 Db 298 GCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 357  
 QY 932 agcattgaccagatcaaacccgtcgtccagtgccagtcag 971  
 Db 358 GTTGTGTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 397

RESULT 4  
 CNS0071A 895 bp DNA GSS 03-JUN-1999  
 LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC #  
 DEFINITION BACR14B09 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.  
 ACCESSION AL066286  
 VERSION AL066286.1 GI:4945153  
 KEYWORDS GSS.  
 SOURCE fruit fly.  
 ORGANISM Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 1 (bases 1 to 895)  
 Genoscope.  
 Direct Submission  
 Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : [seqrefgenoscope.cns.fr](mailto:seqrefgenoscope.cns.fr))  
 - Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr))  
 COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila

melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

FEATURES  
 source Location/Qualifiers  
 1. .895  
 /organism="Drosophila melanogaster"  
 /db\_xref="taxon:7227"  
 /clone\_lib="RPCI-98"  
 /clone="BACR14B09"  
 /note="end : TET3"

BASE COUNT 124 a 80 c 204 g 179 t 308 others  
 ORIGIN

Query Match 2.7%; Score 41.6; DB 13; Length 895;  
 Best Local Similarity 19.9%; Pred. No. 2.7;  
 Matches 68; Conservative 115; Mismatches 159; Indels 0; Gaps 0;

QY 580 ggggtcggcatgtttggcagtggtttcttcgcggaagagtgaaagctcaaggcccttgc 639  
 Db 512 GKGKKGKKGKKGKKGKKGKKGKKGKKGKKGKKGKKGKKGKKGKKGKKGKKGKKG 571  
 QY 640 aggtgcttcaggtgctccgctgttcttggtattacggagcctgcgcttctcggtgtgaa 699  
 Db 572 KKTGKTKTKTKGKTKGKTKGKTKGKTKGKTKGKTKGKTKGKTKGKTKGKTKG 631  
 QY 700 ccttcgctgcgtgcgccttcttcacggtatcggtacgcagctatcggtgcgcttt 759  
 Db 632 KGKTKTKTKGKTKGKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTK 691  
 QY 760 gattgcactctttaaatacaaggcagttgcgttggcgctgcaggtttcttgggtttgt 819  
 Db 692 KKTGTTTGTGTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 751  
 QY 820 ttctattgattccagatatggtcattcttgggtgtgctgagttgttaccttctcat 879  
 Db 752 KGKTK 811  
 QY 880 cgcattcggcgagcagattgcttatggcctttacttctggttcg 921  
 Db 812 TTTGTGKGGKGGKKGKTKKKGKKGKTKTKKKGKTKKKGKTKKKGKTKKKGKTK 853

RESULT 5  
 BE997958 477 bp mRNA EST 06-OCT-2000  
 LOCUS EST429681 GVSU Medicago truncatula cDNA clone pcVSN-8B9, mRNA sequence.  
 DEFINITION BE997958  
 ACCESSION BE997958  
 VERSION BE997958.1 GI:10698234  
 KEYWORDS EST.  
 SOURCE barrel medic.  
 ORGANISM Medicago truncatula  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.  
 1 (bases 1 to 477)  
 Fedorova,M., Pierson,B.L., Samac,D.A., Gantt,J.S., Vance,C.P., Town,C.D., Bowman,C.L., Craven,M.B., Cho,J., Fraser,C.M.  
 ESTs from senescent nodules of Medicago truncatula  
 UNPUBLISHED (2000)  
 JOURNAL Contact: Carroll P. Vance  
 COMMENT Department of Agronomy and Plant Genetics  
 University of Minnesota  
 411 Borlaug Hall, 1991 Upper Buford Circle, St. Paul, MN 55108 USA



REFERENCE 1 (bases 1 to 601)  
 AUTHORS Fedorova,M., Pierson,B.L., Samac,D.A., Vance,C.P., Gantt,G.S., Town  
 C.D., Van Aken,S., Utterback,T., Cho,J. and Fraser,C.M.  
 TITLE ESTs from one month old nitrogen-fixing root nodules of Medicago  
 truncatula, 2001  
 JOURNAL Unpublished (2001)  
 COMMENT Contact: Carroll P. Vance  
 Department of Agronomy and Plant Genetics  
 University of Minnesota  
 411 Borlaug Hall, 1991 Upper Buford Circle, St.Paul, MN 55108 USA  
 Tel: 612 625 5715  
 Fax: 651-649-5058  
 Email: vance004@maroon.tc.umn.edu  
 University of Minnesota name: M381580e TIGR sequence name:  
 MTCT40TK More information is available at: http://www.medicago.org  
 Seq primer: SKmod (CTA gAA CTA gTg gAT CC).

FEATURES  
 Location/Qualifiers  
 1..601  
 /organism="Medicago truncatula"  
 /cultivar="genotype Al7"  
 /db\_xref="taxon:3880"  
 /clone="pGVN-57H8"  
 /clone\_lib="GVN"  
 /tissue\_type="N2-fixing root nodules"  
 /dev\_stage="effective root nodules harvested one month  
 post inoculation with Sinorhizobium meliloti"  
 /lab\_host="E. coli strain XL0LR"  
 /note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2:  
 XhoI; cDNA was prepared from polyA+ enriched RNA from  
 effective root nodules harvested one month post  
 inoculation with Sinorhizobium meliloti. The cDNA was  
 directionally ligated into the Uni-ZAP XR vector from  
 Stragene and packaged using Gigapack III Gold packaging  
 extracts. Plasmids containing cDNA inserts were excised  
 from the recombinant lambda-ZAP phage using Ex-Assist  
 helper phage and propagated in XL0LR cells."  
 BASE COUNT 170 a 121 c 127 g 183 t  
 ORIGIN

Query Match 2.7%; Score 41.4; DB 11; Length 601;  
 Best Local Similarity 50.2%; Pred. No. 2.7;  
 Matches 102; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

QY 1246 cattgtttcgacacagtaaacctcaacgagcagcacttaaccgctgaagaagcagg 1305

Db 307 CATTGACTTCACCAAGAAGAGAGCTTCTCATAGCTTTCTTAAGCCCTAACACGGATCATAA 366

QY 1306 cgatgaagtcaaacgagggagctgtgtgaattcgatattgatgccattaggctgc 1365

Db 367 TGGTTTGTCACTGCATCTGAGTTGCATTATTACTTGACAAATCAAGGATTAAGCGGAC 426

QY 1366 aggttatgaggttaaccacccgattgtgttttcgaattacaagaacccgacctgttaa 1425

Db 427 CAATGAAGAAGTGAGCGACTTTTGTCTGAGGCTGTATTCTGATAGTGATGGACATCTAAG 486

QY 1426 cacttacggtttggcggaattg 1448

Db 487 CTTTAAGGAGTTGTGCAGACTTG 509

RESULT 8  
 BE998413 627 bp mRNA EST 06-OCT-2000  
 LOCUS EST430136 GVSN Medicago truncatula cDNA clone pGVSN-9J12, mRNA  
 DEFINITION sequence.

ACCESSION BE998413  
 VERSION BE998413.1 GI:10698689  
 KEYWORDS EST.  
 SOURCE barrel medic.  
 ORGANISM Medicago truncatula  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;  
 Medicago.  
 1 (bases 1 to 627)  
 AUTHORS Fedorova,M., Pierson,B.L., Samac,D.A., Gantt,J.S., Vance,C.P., Town  
 C.D., Bowman,C.L., Craven,M.B., Cho,J. and Fraser,C.M.  
 TITLE ESTs from senescent nodules of Medicago truncatula  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Carroll P. Vance  
 Department of Agronomy and Plant Genetics  
 University of Minnesota  
 411 Borlaug Hall, 1991 Upper Buford Circle, St.Paul, MN 55108 USA  
 Tel: 612 625 5715  
 Fax: 651-649-5058  
 Email: vance004@maroon.tc.umn.edu  
 University of Minnesota name: M272391e TIGR sequence name:  
 MTKAP54TK More information is available at: http://chrysis.tamu.edu/medicago  
 Seq primer: SKmod (CTA gAA CTA gTg gAT CC).

FEATURES  
 Location/Qualifiers  
 1..627  
 /organism="Medicago truncatula"  
 /cultivar="genotype Al7"  
 /db\_xref="taxon:3880"  
 /clone="pGVSN-9J12"  
 /clone\_lib="GVSN"  
 /tissue\_type="senescent root nodules"  
 /dev\_stage="mixture of effective nodules from 40 day old  
 plants harvested 36 hours post shoot removal and nodules  
 collected from 2 month old plants at mid-pod stage"  
 /lab\_host="E. coli strain SOLR"  
 /note="Vector: pBluescript SK +/-; Site\_1: EcoRI; Site\_2:  
 XhoI; cDNA was prepared from polyA+ enriched RNA from the  
 mixture of effective nodules of 40 day old plants  
 harvested 36 hours post shoot removal and nodules  
 collected from 2 month old plants at mid-pod stage. The  
 cDNA was directionally ligated into the Uni-ZAP XR vector  
 from Stragene and packaged using Gigapack III Gold  
 packaging extracts. Plasmids containing cDNA inserts were  
 excised from the recombinant lambda-ZAP phage using  
 Ex-Assist helper phage and propagated in SOLR cells."  
 BASE COUNT 175 a 124 c 135 g 193 t  
 ORIGIN

Query Match 2.7%; Score 41.4; DB 11; Length 627;  
 Best Local Similarity 50.2%; Pred. No. 2.8;  
 Matches 102; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

QY 1246 cattgtttcgacacagtaaacctcaacgagcagcacttaaccgctgaagaagcagg 1305

Db 313 CATTGACTTCACCAAGAAGAGAGCTTCTCATAGCTTTCTTAAGCCCTAACACGGATCATAA 372

QY 1306 cgatgaagtcaaacgagggagctgtgtgaattcgatattgatgccattaggctgc 1365

Db 373 TGGTTTGTCACTGCATCTGAGTTGCATTATTACTTGACAAATCAAGGATTAAGCGGAC 432

QY 1366 aggttatgaggttaaccacccgattgtgttttcgaattacaagaacccgacctgttaa 1425

Db 433 CAATGAAGAAGTGAGCGACTTTTGTCTGAGGCTGTATTCTGATAGTGATGGACATCTAAG 492

QY 1426 cacttacggtttggcggaattg 1448

Db 493 CTTTAAGGAGTTGTGCAGACTTG 515

RESULT 9  
 CNS006U0 884 bp DNA GSS 03-JUN-1999  
 LOCUS Drosophila melanogaster genome survey sequence T7 end of BAC #  
 DEFINITION BAC14N21 of RPCI-98 library from Drosophila melanogaster (fruit  
 fly), genomic survey sequence.  
 ACCESSION AL065923  
 VERSION AL065923.1 GI:4944891

```

KEYWORDS      GSS.
SOURCE         fruit fly.
ORGANISM       Drosophila melanogaster
               Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
               Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
               Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE      1 (bases 1 to 884)
AUTHORS        Direct Submission
TITLE          Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
JOURNAL        BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
               - Web : www.genoscope.cns.fr)
COMMENT        Determination of this BAC-end
               collaboration with the Berkeley Drosophila Genome Project (BDGP).
               The BDGP is constructing a physical map of the Drosophila
               melanogaster genome using these BACs. For further information
               please see http://www.fruitfly.org The BDGP Drosophila
               melanogaster BAC library was prepared by Kazutoyo Osoegawa and
               Aaron Mamoser in Pieter de Jong's laboratory in the Department of
               Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
               NY. The library is named RPCI-98 and was constructed by partial
               EcoRI digestion of Drosophila DNA provided by the BDGP from the
               isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
               PI and EST libraries. A more detailed description of the library
               and how to order individual BAC clones, the entire library, or
               filters for hybridization from the BACPAC Resource Center can be
               found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES       Location/Qualifiers
               1..884
               /organism="Drosophila melanogaster"
               /db_xref="taxon:7227"
               /clone_lib="RPCI-98"
               /clone="BACR14N21"
               /note="end : 77"

BASE COUNT     230 a   62 c   139 g   124 t   329 others
ORIGIN

Query Match    2.7%; Score 40.8; DB 13; Length 884;
Best Local Similarity 15.0%; Pred. No. 4.4;
Matches 55; Conservative 147; Mismatches 165; Indels 0; Gaps 0;

QY 572 atcgccaggggtgagcattgttggcagtggtttcttcctggcgaagagtgaaagctcaag 631
Db 516 RKGVGMVAVKTTGGGGGKRTGTTKGGKSTKGDGKGVGTGKTKTTKDDTTGKKTGG 575

QY 632 ggccttcagagtgcttcagtggtctcgcgtgttcttggattacggagcctgcgatcttc 691
Db 576 TGAGKTGGGKGGGTGGGTGGGKGGKGGTGTGTTKTKTKTKTKTKTKTKTKTKTKTKTK 635

QY 692 ggtgtgaaccttcgcctgcgctgttcttcacgtatcggtaccgcagctatcggt 751
Db 636 GGGGGGGKKGKGGTGKGGGGKKGKGTGTRGDRDGTGKKTGKTKKTKKTKKTKGKG 695

QY 752 ggcgcttgattgcactctttaatacaaggcagttgcgttggcgctgcaggtttcttc 811
Db 696 KGKTKKKSgtKKGKGGTGKGTGKGTGKTKKKGKGGKGGKGGKKGKKGKKGKKGK 755

QY 812 ggtgtgtttcttattgatgctccagatattgctatgttcttcttgggtggtgagttgacc 871
Db 756 KKKKGGKKGKTKGKGGTKGKGGTKGKGGTKGKGGTKGKGGTKGKGGTKGKGGTKGKGG 815

QY 872 ttcttcacgcttcgagcagcagtgcttgcattggtcttacttgcgttcgagcagcgc 931
Db 816 KKKKKKKTKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK 875

QY 932 agcattg 938
Db 876 KKBKKGK 882

RESULT 10
CNS03CVD/c

```

```

LOCUS          CNS03CVD      983 bp      DNA      GSS      15-MAY-2000
DEFINITION     Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
                015D07 of library G from Tetraodon nigroviridis, genomic survey
                sequence.
ACCESSION      AL238306
VERSION        AL238306.1 GI:7897441
KEYWORDS       GSS; genome survey sequence.
SOURCE         Tetraodon nigroviridis.
ORGANISM       Tetraodon nigroviridis.
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
               Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
               Tetraodontidae; Tetraodon.
REFERENCE      1 (bases 1 to 983)
AUTHORS        Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
               Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
               Weissenbach,J.
TITLE          Characterization and repeat analysis of the compact genome of the
               freshwater pufferfish Tetraodon nigroviridis
JOURNAL        Unpublished
REFERENCE      2 (bases 1 to 983)
AUTHORS        Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
               Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
               Saurin,W. and Weissenbach,J.
TITLE          Human gene number estimate provided by genome wide analysis using
               Tetraodon nigroviridis DNA sequence
JOURNAL        Unpublished
REFERENCE      3 (bases 1 to 983)
AUTHORS        Genoscope.
TITLE          Direct Submission
JOURNAL        Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
COMMENT        This sequence is a single read and was generated as part of a large
               scale clone-end sequencing project of the Tetraodon nigroviridis
               genome. For more information, please take a look at
               http://www.genoscope.cns.fr/Tetraodon.

FEATURES       Location/Qualifiers
               1..983
               /organism="Tetraodon nigroviridis"
               /db_xref="taxon:99883"
               /clone="015D07"
               /clone_lib="G"
               /note="Genoscope sequence ID : C0BG015CB04SP1-end :
               PUC-Ori"

BASE COUNT     202 a   289 c   274 g   201 t   17 others
ORIGIN

Query Match    2.7%; Score 40.8; DB 13; Length 983;
Best Local Similarity 53.0%; Pred. No. 4.5;
Matches 87; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY 705 gcttcgctgctgcttcttcacgtatcggtaccgcagctatcggtgctgcttgcattg 764
Db 732 GCCGGAGCCGGGAGCTGGCCCCCGCTCCGCCGCGCTGGGGGTTTTATTACCGT 673

QY 765 cactcttaatacaaggcagttgcgttggcgctgcaggttcttcttgggtgtgttctcta 824
Db 672 TGTATTATTATCATATTATTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 613

QY 825 ttgatctccagatattgctatgttcttcttgggtggtgctgagttgtt 868
Db 612 TTGTTGCGTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 569

RESULT 11
CNS00LXJ
LOCUS          CNS00LXJ      1101 bp      DNA      GSS      14-JUN-1999
DEFINITION     Drosophila melanogaster genome survey sequence T7 end of BAC:
                BACR48E16 of RPCI-98 library from Drosophila melanogaster (fruit
                fly), genomic survey sequence.
ACCESSION      AL078875
VERSION        AL078875.1 GI:5102165
KEYWORDS       GSS.

```



KEYWORDS EST.  
SOURCE barrel medic.  
ORGANISM Medicago truncatula

REFERENCE 1 (bases 1 to 606)  
AUTHORS Fedorova, M., Pierson, B.L., Samac, D.A., Gantt, J.S., Vance, C.P., Town  
C.D., Bowman, C.L., Craven, M.B., Cho, J., and Fraser, C.M.  
TITLE ESTs from senescent nodules of Medicago truncatula  
JOURNAL Unpublished (2000)  
CONTACT: Carroll P. Vance  
Department of Agronomy and Plant Genetics  
University of Minnesota  
411 Borlaug Hall, 1991 Upper Buford Circle, St. Paul, MN 55108 USA  
Tel: 612 625 5715  
Fax: 651-649-5058  
Email: vance004@maroon.tc.umn.edu  
University of Minnesota name: M271935e TIGR sequence name:  
MTKAK05NK More information is available at:  
http://chrysie.tamu.edu/medicago  
Seq primer: Skmcd (CTA gAA CTA gtg gAT CC).

FEATURES  
source  
1..606  
/organism="Medicago truncatula"  
/cultivar="genotype A17"  
/db\_xref="taxon:3880"  
/clone="pGVSN-889"  
/clone\_lib="GVSN"  
/tissue\_type="senescent root nodules"  
/dev\_stage="mixture of effective nodules from 40 day old  
plants harvested 36 hours post shoot removal and nodules  
collected from 2 month old plants at mid-pod stage"  
/lab\_host="E. coli strain SOLR"  
/note="vector: pBluescript SK +/-; Site\_1: EcoRI; Site\_2:  
XhoI; cDNA was prepared from polyA+ enriched RNA from the  
mixture of effective nodules of 40 day old plants  
harvested 36 hours post shoot removal and nodules  
collected from 2 month old plants at mid-pod stage. The  
cDNA was directionally ligated into the Uni-ZAP XR vector  
from Stratagene and packaged using Gigapack III Gold  
packaging extracts. Plasmids containing cDNA inserts were  
excised from the recombinant lambda-ZAP phage using  
Ex-Assist helper phage and propagated in SOLR cells."

BASE COUNT 173 a 122 c 127 g 184 t  
ORIGIN

Query Match 2.6%; Score 39.8; DB 11; Length 606;  
Best Local Similarity 49.8%; Pred. No. 7.3;  
Matches 101; Conservative 0; Mismatches 102; Indels 0; Gaps 0;

QY 1246 cattggtttcacacagtaaacctcaacgcgcagcactttaaccctgaagcagg 1305  
Db 317 CATTGACTTACCACAAAGAGAGCTTCTCATAGCTTCTCTAAGCCTTAACACGGATCATAA 376  
QY 1306 cgaagaagcaaacaggaggagctgtgtgaattcgatattgatgccattaagctgc 1365  
Db 377 TGGTTTGTGCATGTCATGTGAGTTGATTACTTGCACAAATCAAGGCATTAAGCGCAC 436  
QY 1366 aggttatgaggttaacacgcgcgattgttttcgaattacaagaaacccggacctgtaaa 1425  
Db 437 CAATGAAGAGTGAGCGCACTTTCTTCTGAGGCTGATTCTGATAGTGGACATCTAAG 496

QY 1426 cacttacggtttggcgcaattg 1448  
Db 497 CTTTAAAGAGCTTGTTCAGACTTG 519

RESULT 14  
CNS0010P/c 993 bp DNA GSS 03-JUN-1999  
LOCUS

DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC:  
BACR23F02 of RPCI-98 library from Drosophila melanogaster (fruit  
fly), genomic survey sequence.  
ACCESSION AL067821  
VERSION AL067821.1 GI:4957842  
KEYWORDS GSS.  
SOURCE fruit fly.  
ORGANISM Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
REFERENCE 1 (bases 1 to 993)  
AUTHORS Direct Submission  
TITLE Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :  
JOURNAL BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
COMMENT - Web : www.genoscope.cns.fr)  
Determination of this BAC-end sequence was carried out as part of a  
collaboration with the Berkeley Drosophila Genome Project (BDGP).  
The BDGP is constructing a physical map of the Drosophila  
melanogaster genome using these BACs. For further information  
please see <http://www.fruitfly.org/TheBDGP/Drosophila>  
melanogaster BAC library was prepared by Kazutoyo Osoegawa and  
Aaron Mamoser in Pieter de Jong's laboratory in the Department of  
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,  
NY. The library is named RPCI-98 and was constructed by partial  
EcoRI digestion of Drosophila DNA provided by the BDGP from the  
isogenic strain y2; cn bw sp, the same strain used for the BDGP's  
P1 and EST libraries. A more detailed description of the library  
and how to order individual BAC clones, the entire library, or  
filters for hybridization from the BACPAC Resource Center can be  
found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

FEATURES  
source  
1..993  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/clone\_lib="RPCI-98"  
/clone="BACR23F02"  
/note="end : T7"

BASE COUNT 379 a 207 c 186 g 151 t 70 others  
ORIGIN

Query Match 2.6%; Score 39.2; DB 13; Length 993;  
Best Local Similarity 48.4%; Pred. No. 12;  
Matches 104; Conservative 1; Mismatches 110; Indels 0; Gaps 0;

QY 705 gctgcgctggccgtttcttcacgttatcggtaccgcagctatcggtggcgctttgattg 764  
Db 229 GCTGCGGTTTGTGGCGCTCCATTGATGCTGCTGCTGCTGCTGGAGCTGCTCTGTTG 170  
QY 765 cactctttaatatcaaggcagttgcgttggcgctgcaggtttcttgggtgtgtttcta 824  
Db 169 CTGTTGCTGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 110  
QY 825 ttgatctccagatatggctcatgttcttgggtgtgagtcagttgttaccttcttcacat 884  
Db 109 TTGATGTTGCTGCTGCTGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 50  
QY 885 tggcgccagcagttgattggcctttacttgggt 919  
Db 49 TTGTTGGCGCTGTTGCTTGTGTTTTTTTCTCTTTTTT 15

RESULT 15  
AI070146/c 539 bp mRNA EST 05-JUL-1999  
LOCUS AI070146  
DEFINITION UI-R-Y0-lu-g-12-0-UI-s1 UI-R-Y0 Rattus norvegicus cDNA clone  
UI-R-Y0-lu-g-12-0-UI 3', mRNA sequence.  
ACCESSION AI070146  
VERSION AI070146.1 GI:3396397  
KEYWORDS EST.  
SOURCE Norway rat.

ORGANISM	Rattus norvegicus Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE	1 (bases 1 to 539)
AUTHORS	Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE	Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL	Genome Res. 6 (9), 791-806 (1996)
MEDLINE	9704477
COMMENT	Contact: Soares, MB Program for Rat Gene Discovery and Mapping University of Iowa 451 Eckstein Medical Research Building Iowa City, IA 52242, USA Tel: 319 335 8250 Fax: 319 335 9565

Email: msoares@blue.weeg.uiowa.edu

The sequence tag present in the cDNA between the NotI site and the Oligo-dT track served to identify it as a clone from the normalized adult Eye library. cDNA Library Preparation: M. Fatima Bonaldo, Ph.D. Clone distribution: clones will be available through Research Genetics This clone is also available through the I.M.A.G.E. Consortium at LLM (info@image.llnl.gov). IMAGE ID=195384 The following repetitive elements were found in this cDNA sequence: 357-476, >(CAA)n#Simple\_repeat 478-538, >(CAA)n#Simple\_repeat seq primer: M13 Forward

FEATURES	SOURCE
1. <b>Feature 1</b>	Source 1
2. <b>Feature 2</b>	Source 2
3. <b>Feature 3</b>	Source 3
4. <b>Feature 4</b>	Source 4
5. <b>Feature 5</b>	Source 5
6. <b>Feature 6</b>	Source 6
7. <b>Feature 7</b>	Source 7
8. <b>Feature 8</b>	Source 8
9. <b>Feature 9</b>	Source 9
10. <b>Feature 10</b>	Source 10
11. <b>Feature 11</b>	Source 11
12. <b>Feature 12</b>	Source 12
13. <b>Feature 13</b>	Source 13
14. <b>Feature 14</b>	Source 14
15. <b>Feature 15</b>	Source 15
16. <b>Feature 16</b>	Source 16
17. <b>Feature 17</b>	Source 17
18. <b>Feature 18</b>	Source 18
19. <b>Feature 19</b>	Source 19
20. <b>Feature 20</b>	Source 20
21. <b>Feature 21</b>	Source 21
22. <b>Feature 22</b>	Source 22
23. <b>Feature 23</b>	Source 23
24. <b>Feature 24</b>	Source 24
25. <b>Feature 25</b>	Source 25
26. <b>Feature 26</b>	Source 26
27. <b>Feature 27</b>	Source 27
28. <b>Feature 28</b>	Source 28
29. <b>Feature 29</b>	Source 29
30. <b>Feature 30</b>	Source 30
31. <b>Feature 31</b>	Source 31
32. <b>Feature 32</b>	Source 32
33. <b>Feature 33</b>	Source 33
34. <b>Feature 34</b>	Source 34
35. <b>Feature 35</b>	Source 35
36. <b>Feature 36</b>	Source 36
37. <b>Feature 37</b>	Source 37
38. <b>Feature 38</b>	Source 38
39. <b>Feature 39</b>	Source 39
40. <b>Feature 40</b>	Source 40
41. <b>Feature 41</b>	Source 41
42. <b>Feature 42</b>	Source 42
43. <b>Feature 43</b>	Source 43
44. <b>Feature 44</b>	Source 44
45. <b>Feature 45</b>	Source 45
46. <b>Feature 46</b>	Source 46
47. <b>Feature 47</b>	Source 47
48. <b>Feature 48</b>	Source 48
49. <b>Feature 49</b>	Source 49
50. <b>Feature 50</b>	Source 50
51. <b>Feature 51</b>	Source 51
52. <b>Feature 52</b>	Source 52
53. <b>Feature 53</b>	Source 53
54. <b>Feature 54</b>	Source 54
55. <b>Feature 55</b>	Source 55
56. <b>Feature 56</b>	Source 56
57. <b>Feature 57</b>	Source 57
58. <b>Feature 58</b>	Source 58
59. <b>Feature 59</b>	Source 59
60. <b>Feature 60</b>	Source 60
61. <b>Feature 61</b>	Source 61
62. <b>Feature 62</b>	Source 62
63. <b>Feature 63</b>	Source 63
64. <b>Feature 64</b>	Source 64
65. <b>Feature 65</b>	Source 65
66. <b>Feature 66</b>	Source 66
67. <b>Feature 67</b>	Source 67
68. <b>Feature 68</b>	Source 68
69. <b>Feature 69</b>	Source 69
70. <b>Feature 70</b>	Source 70
71. <b>Feature 71</b>	Source 71
72. <b>Feature 72</b>	Source 72
73. <b>Feature 73</b>	Source 73
74. <b>Feature 74</b>	Source 74
75. <b>Feature 75</b>	Source 75
76. <b>Feature 76</b>	Source 76
77. <b>Feature 77</b>	Source 77
78. <b>Feature 78</b>	Source 78
79. <b>Feature 79</b>	Source 79
80. <b>Feature 80</b>	Source 80
81. <b>Feature 81</b>	Source 81
82. <b>Feature 82</b>	Source 82
83. <b>Feature 83</b>	Source 83
84. <b>Feature 84</b>	Source 84
85. <b>Feature 85</b>	Source 85
86. <b>Feature 86</b>	Source 86
87. <b>Feature 87</b>	Source 87
88. <b>Feature 88</b>	Source 88
89. <b>Feature 89</b>	Source 89
90. <b>Feature 90</b>	Source 90
91. <b>Feature 91</b>	Source 91
92. <b>Feature 92</b>	Source 92
93. <b>Feature 93</b>	Source 93
94. <b>Feature 94</b>	Source 94
95. <b>Feature 95</b>	Source 95
96. <b>Feature 96</b>	Source 96
97. <b>Feature 97</b>	Source 97
98. <b>Feature 98</b>	Source 98
99. <b>Feature 99</b>	Source 99
100. <b>Feature 100</b>	Source 100

**Location/Qualifiers**

1. 539

/location/quadri1ers

/organism="Rattus norvegicus"

/strain="Sprague-Dawley"

/db\_xref="taxon:10116"

/clone="UI-R-Y0-lu-g-12-0-UI"

/clone\_l1b="UI-R-Y0"

/dev\_stage="adult"

/lab\_host="DH10B (Life Technologies)"

/note="vector: p7T30-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; The UI-R-Y0 library is a subtracted library derived from an individually-tagged normalized whole-eye (minus the lens) library. The driver for the subtraction consisted of a pool of all previous libraries (UI-R-A0, UI-R-A1, UI-R-E0, UI-R-E1, UI-R-C0 and UI-R-C1). The tag is a string of 3-5 nucleotides present between the Not I site and the oligo-dT track which allows identification of the library of origin of a clone within the mixture. The subtracted library (UI-R-Y0) was constructed as follows: PCR amplified cDNA inserts from previous library clones from which 3' ESTs had been derived were used as a driver in a hybridization with the normalized whole-eye library in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (Life Technologies) to generate the UI-R-Y0 library. This procedure has been previously described (Bonaldi, Lennon and Soares, Genome Research 6: 791-806, 1996)"

1990			
BASE COUNT	202 a	117 c	92 g
ORIGIN			
			128 t

Query Match 2.5%; Score 38.8; DB 10; Length 539;  
Best Local Similarity 51.1%; Pred. No. 13;  
Matches 91; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

Qy	726	tcggtatcggtaaccgcagcgtatcggtggcgctttgattggcactctttaaatacaaggcag	785
Db	521	ttgtttttctgtttgctgtgtgtattttctgtttgtttgtttgtttgtgtgtgt	462
Qy	786	tgcgctggggcgtcgaggttctctgggtgttttctcatgatctccagatataggcca	845
Db	461	ttgtgtttatttcgtgtgtgtttttgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	402

**Qy** 846 t g t t c t t g g t g t g c a g t t g t t a c c t t c t c a t t c g c a t t c g g c a g c g a t t g c t t a 903  
| | | | | | | | | | | | | | | | | | | | | |  
**Db** 401 t t t t t g t t g t g t g c t g c t t g t t a t t t c t g t t g t t g t t g t t g c t g c t g c t g t t t a 344

Search completed: March 22, 2002, 08:11:36  
Job time: 5480 sec

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 22, 2002, 09:23:24 ; Search time 66.97 Seconds  
(without alignments)  
3750.397 Million cell updates/sec

Title: US-09-604-231-3

Perfect score: 1109

Sequence: 1 tatgatttcggcggtccagt.....gttgaacattgagttctcg 1109

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents, NA:\*

- 1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:\*
- 2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:\*
- 3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq:\*
- 4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq:\*
- 5: /cgn2\_6/ptodata/2/ina/PCFUS\_COMB.seq:\*
- 6: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	91	8.2	3615	1	US-08-920-812-17
C 2	91	8.2	3615	1	US-08-920-827-17
C 3	91	8.2	3615	1	US-08-921-177-17
C 4	91	8.2	3615	1	US-08-362-577C-17
C 5	91	8.2	3615	2	US-08-920-828-17
C 6	39.8	3.6	465	2	US-08-673-190A-3
C 7	39.6	3.6	7218	1	US-08-232-463-14
C 8	37.2	3.4	4403765	4	US-09-103-840A-2
C 9	32.6	2.9	477	4	US-09-135-994-1
C 10	32.4	2.9	292	4	US-09-117-121-29
C 11	32.2	2.9	7218	1	US-08-232-463-14
C 12	32	2.9	1008	4	US-09-198-955A-9
C 13	32	2.9	3533	4	US-08-985-916-15
C 14	32	2.9	5163	3	US-08-700-651-1
C 15	32	2.9	5163	3	US-08-928-361B-4
C 16	32	2.9	5318	3	US-08-700-651-2
C 17	32	2.9	5318	3	US-08-928-361B-3
C 18	32	2.9	10348	2	US-08-457-273B-41
C 19	32	2.9	10348	3	US-08-556-419-13
C 20	32	2.9	10348	4	US-09-041-886-14
C 21	32	2.9	10366	1	US-08-246-982A-5
C 22	32	2.9	10366	1	US-08-453-285-5
C 23	31.6	2.8	5511	3	US-08-928-361B-2
C 24	31.6	2.8	7334	3	US-08-928-361B-1
C 25	31.6	2.8	7655	1	US-08-619-554-1
C 26	31.2	2.8	342	4	US-08-905-223-196
C 27	30.8	2.8	2745	1	US-08-363-255-1

28	30.8	2.8	2745	1	US-08-363-255-13	Sequence 13, Appl
C 29	30.6	2.8	289	4	US-09-007-005-17	Sequence 17, Appl
C 30	30.6	2.8	289	4	US-09-244-796-17	Sequence 17, Appl
C 31	30.4	2.7	688	4	US-08-998-416-915	Sequence 915, App
C 32	30.4	2.7	3807	1	US-08-022-835-5	Sequence 5, Appl
C 33	30.4	2.7	3807	1	US-08-388-809-5	Sequence 5, Appl
C 34	30.4	2.7	3807	2	US-08-647-714-5	Sequence 5, Appl
C 35	30.4	2.7	20303	1	US-08-370-975B-6	Sequence 6, Appl
C 36	30.4	2.7	26764	1	US-08-370-975B-1	Sequence 1, Appl
C 37	30.4	2.7	19136	4	US-09-422-869-1	Sequence 1, Appl
C 38	30.2	2.7	1996	2	US-08-559-524A-1	Sequence 1, Appl
C 39	30.2	2.7	1996	3	US-08-749-707-1	Sequence 2, Appl
C 40	30.2	2.7	2277	1	US-08-676-967-2	Sequence 2, Appl
C 41	30.2	2.7	2277	1	US-08-676-974-2	Sequence 2, Appl
C 42	30.2	2.7	2277	2	US-09-098-487-2	Sequence 2, Appl
C 43	30.2	2.7	4692	2	US-08-916-917-1	Sequence 1, Appl
C 44	30.2	2.7	4692	2	US-08-972-631-1	Sequence 1, Appl
C 45	30.2	2.7	4692	2	US-08-972-629-1	Sequence 1, Appl

#### ALIGNMENTS

RESULT 1  
US-08-920-812-17/c  
; Sequence 17, Application US/08920812  
; Patent No. 5763188  
; GENERAL INFORMATION:  
; APPLICANT: Ohno, Tsuneoya  
; APPLICANT: Matsuhisa, Akio  
; APPLICANT: Uehara, Hirotugu  
; APPLICANT: Eda, Soji  
; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease  
; NUMBER OF SEQUENCES: 25  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 6300 Sears Tower, 233 South Wacker Drive  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States of America  
; ZIP: 60606-6402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/920,812  
; FILING DATE: 29-AUG-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/362,577  
; FILING DATE: 27-MAR-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Rin-Laures, Li-Hsien  
; REGISTRATION NUMBER: 33,547  
; REFERENCE/DOCKET NUMBER: 19036/32420  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312/474-6300  
; TELEFAX: 312/474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 17:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3615 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: Genomic DNA  
; ORIGINAL SOURCE:  
; ORGANISM: Escherichia coli  
; STRAIN: Clinical Isolate EC-24  
US-08-920-812-17





```

Db 1450 GTATTATTCAGTAAATACCGATGATTTTACGGAGCTGATTATACCCACCGCAGCGGCGAGATA 139
QY 1030 gaagcgaggagccaaacctgctcaacgtc 1056
    ||| ||| ||| ||| ||| |||
Db 1390 AGCGCAGGTGAACCGCGTGTTATCCATC 1364

RESULT 5
US-08-920-828-17/c
: Sequence 17, Application US/08920828
: Patent No. 5853998
: GENERAL INFORMATION:
: APPLICANT: Ohno, Tsuneya
: APPLICANT: Matsuhisa, Akio
: APPLICANT: Uehara, Hirotugu
: APPLICANT: Eda, Soji
: TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
: NUMBER OF SEQUENCES: 25
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
: STREET: 6300 Sears Tower, 233 South Wacker Drive
: CITY: Chicago
: STATE: Illinois
: COUNTRY: United States of America
: ZIP: 60606-6402
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/920,828
: FILING DATE: 29-AUG-1997
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/362,577
: FILING DATE: 27-MAR-1995
: ATTORNEY/AGENT INFORMATION:
: NAME: Rin-Laures, Li-Hsien
: REGISTRATION NUMBER: 33,547
: REFERENCE/DOCKET NUMBER: 19036/32420
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 312/474-6300
: TELEFAX: 312/474-0448
: TELEX: 25-3856
: INFORMATION FOR SEQ ID NO: 17:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3615 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: Genomic DNA
: ORIGINAL SOURCE:
: ORGANISM: Escherichia coli
: STRAIN: Clinical Isolate EC-24
US-08-920-828-17

Query Match 8.2%; Score 91; DB 2; Length 3615;
Best Local Similarity 53.7%; Pred. No. 6.4e-19;
Matches 240; Conservative 0; Mismatches 195; Indels 12; Gaps

QY 610 atccaggcacctttgaacggtgaagctattgacctgagcagcgtcagcgatgccatgttt 669
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1798 AUTCCTGTTACCGATGACGGGAGAGATTGTGCGCTCATTCACGTCGCTGATPACCACGTTT 1739
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 670 gccacgagaaagcttggctcgggcgttggccatcgctcccaaccaaggggcagtttagttct 729
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1738 GCCAGTGGCCTGTGTGGTAAAGGATTGCCATTTCGCCCTCGGTTGGTGAAGTCGCTTCT 1679
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 730 ccggtgagtggaagattggtggcatctcccatgtggaatggttcgagtttcgacc 789
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1678 CCGTGTTCGGGGTCGAATTCTTCGTTGTTTCGCCACATTACA-----CGCATTTGCCATT 1625
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```



COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/117,121  
FILING DATE: 20-NOV-1998  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/CA97/00062  
FILING DATE: 30-JAN-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Weber, Kenneth A.  
REGISTRATION NUMBER: 31,677  
REFERENCES/DOCKET NUMBER: 016252-001610US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 29:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 292 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA

2.98: score 32.2; DB 1; Length 7218;







GenCore version 4.5  
 Copyright (c) 1993 - 2000 Compugen Ltd.  
 OM nucleic - nucleic search, using sw model  
 Run on: March 22, 2002, 08:53:35 ; Search time 157.11 Seconds  
 (without alignments)  
 6051.643 Million cell updates/sec

Title: US-09-604-231-3  
 Perfect score: 1109  
 Sequence: 1 tatgatttcggcggtccagt.....gttgaaaccttgatttgcg 1109

Scoring table: IDENTITY\_NUC  
 Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_1101.\*  
 1: /SID52/gcgdata/geneseq/geneseqn/NA1980.DAT.\*  
 2: /SID52/gcgdata/geneseq/geneseqn/NA1981.DAT.\*  
 3: /SID52/gcgdata/geneseq/geneseqn/NA1982.DAT.\*  
 4: /SID52/gcgdata/geneseq/geneseqn/NA1983.DAT.\*  
 5: /SID52/gcgdata/geneseq/geneseqn/NA1984.DAT.\*  
 6: /SID52/gcgdata/geneseq/geneseqn/NA1985.DAT.\*  
 7: /SID52/gcgdata/geneseq/geneseqn/NA1986.DAT.\*  
 8: /SID52/gcgdata/geneseq/geneseqn/NA1987.DAT.\*  
 9: /SID52/gcgdata/geneseq/geneseqn/NA1988.DAT.\*  
 10: /SID52/gcgdata/geneseq/geneseqn/NA1989.DAT.\*  
 11: /SID52/gcgdata/geneseq/geneseqn/NA1990.DAT.\*  
 12: /SID52/gcgdata/geneseq/geneseqn/NA1991.DAT.\*  
 13: /SID52/gcgdata/geneseq/geneseqn/NA1992.DAT.\*  
 14: /SID52/gcgdata/geneseq/geneseqn/NA1993.DAT.\*  
 15: /SID52/gcgdata/geneseq/geneseqn/NA1994.DAT.\*  
 16: /SID52/gcgdata/geneseq/geneseqn/NA1995.DAT.\*  
 17: /SID52/gcgdata/geneseq/geneseqn/NA1996.DAT.\*  
 18: /SID52/gcgdata/geneseq/geneseqn/NA1997.DAT.\*  
 19: /SID52/gcgdata/geneseq/geneseqn/NA1998.DAT.\*  
 20: /SID52/gcgdata/geneseq/geneseqn/NA1999.DAT.\*  
 21: /SID52/gcgdata/geneseq/geneseqn/NA2000.DAT.\*  
 22: /SID52/gcgdata/geneseq/geneseqn/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1109	100.0	1109	22 AAF31529	C-glutamicum phosph
2	1107.4	99.9	1527	22 AAF31528	C-glutamicum phosph
3	1107.4	99.9	349980	22 AAH68533	C glutamicum codin
4	1084.4	97.8	1983	22 AAH67869	C glutamicum codin
5	1049.8	94.7	5969	22 AAF32543	Brevibacterium lac
6	128.6	11.6	5840	20 AAX12968	Enterococcus faeca
7	123.2	11.1	2913	22 AAH54100	S. epidermidis gen
8	118.8	10.7	3895	19 AAV52334	Streptococcus pneu
9	91	8.2	3615	15 AAQ55752	Escherichia coli g
10	90.2	8.1	9769	19 AAV52163	Streptococcus pneu
11	76.4	6.9	465	18 AAV75628	Staphylococcus aur

12	67	6.0	8494	19	AAV52296	Streptococcus pneu
13	66	6.0	1947	22	AAH81338	Escherichia coli p
c 14	65.2	5.9	9797	20	AAX13487	Enterococcus faeca
15	59.2	5.3	2049	22	AAH68426	C glutamicum codin
16	59.2	5.3	2172	22	AAF31542	C glutamicum codin
17	59.2	5.3	349980	22	AAH68528	C glutamicum codin
18	54.2	4.9	428	22	AAF31544	C glutamicum phosph
c 19	53.2	4.8	2996	22	AAH54445	S. epidermidis gen
c 20	53.2	4.8	3081	22	AAH54946	S. epidermidis gen
c 21	53.2	4.8	3932	22	AAH54056	S. epidermidis gen
c 22	51.8	4.7	7156	20	AAX12966	Enterococcus faeca
c 23	51.4	4.6	6285	20	AAX13352	Enterococcus faeca
c 24	51.4	4.6	28882	19	AAV52273	Streptococcus pneu
c 25	51	4.6	29555	18	AAV74517	Staphylococcus aur
c 26	49.2	4.4	796	18	AAV74736	Staphylococcus aur
c 27	46	4.1	1906	20	AAX13595	Enterococcus faeca
c 28	45	4.1	3604	20	AAX13417	Enterococcus faeca
c 29	44.6	4.0	7900	18	AAV74449	Staphylococcus aur
c 30	44.2	4.0	6092	20	AAX13162	Enterococcus faeca
c 31	43.2	3.9	760	19	AAZ96397	S. pneumoniae deri
c 32	43	3.9	29555	18	AAV74517	Staphylococcus aur
c 33	42.8	3.9	30246	18	AAV74367	Staphylococcus aur
c 34	42	3.8	400	18	AAV78285	Staphylococcus aur
c 35	41.6	3.8	341	18	AAV78546	Staphylococcus aur
c 36	40.6	3.7	567	21	AAA29550	HIV codon altered
c 37	40	3.6	249	20	AAX13801	Enterococcus faeca
c 38	39.8	3.6	465	18	AAV74502	Partial P. rubrum
c 39	39.4	3.6	3889	22	AAH54860	S. epidermidis gen
c 40	38.6	3.5	2365	18	AAV74406	Staphylococcus aur
c 41	37.6	3.4	10732	21	AAA10594	Gene encoding a su
c 42	37.4	3.4	5059	21	AAZ36227	DNA encoding a per
c 43	37.2	3.4	65632	21	AAH81502	N. meningitidis pa
c 44	37.2	3.4	349980	21	AAF21544	Neisseria meningit
45	37.2	3.4	1437668	21	AAA81490	N. meningitidis B

## ALIGNMENTS

## RESULT 1

AAF31529  
 ID AAF31529 standard; DNA; 1109 BP.  
 XX  
 AC AAF31529;  
 XX  
 DT 09-APR-2001 (first entry)  
 XX  
 DE C-glutamicum phosphoenolpyruvate DNA #2.  
 XX  
 KW Phosphoenolpyruvate; sugar phosphotransferase system; PTS; ds.  
 XX  
 OS Corynebacterium glutamicum.  
 XX  
 PN WO200102583-A2.  
 XX  
 PD 11-JAN-2001.  
 XX  
 PF 27-JUN-2000; 2000WO-IB00973.  
 XX  
 PR 01-JUL-1999; 99US-0142691.  
 PR 23-AUG-1999; 99US-0150310.  
 PR 03-SEP-1999; 99DE-1042095.  
 PR 03-SEP-1999; 99DE-1042097.  
 (BADI ) BASF AG.  
 Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;  
 WPI; 2001-080989/09.  
 Corynebacterium glutamicum nucleic acids encoding phosphoenolpyruvate:  
 sugar phosphotransferase system proteins or their portions, useful for  
 typing or identifying C. glutamicum or related bacteria, and as markers



Query Match 99.9%; Score 1107.4; DB 22; Length 1527;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1108; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 tatgatttcggtccagtcggtctgctctcggtctgctgtctgtctactcaccatcgctc 60  
Db 419 tatgatttcggtccagtcggtctgctctcggtctgctgtctgtctactcaccatcgctc 478

Qy 61 atcaatggtctgcacacagtccttcgcgaattgagctgagctgttttaaccaggtgga 120  
Db 479 atcaatggtctgcacacagtccttcgcgaattgagctgagctgttttaaccaggtgga 538

Qy 121 tcttcacatctgcgaacggcatctatgctaataatcgcccaagggtgcggcatgtttggca 180  
Db 539 tcttcacatctgcgaacggcatctatgctaataatcgcccaagggtgcggcatgtttggca 598

Qy 181 gtgtcttccttcggtggaagagtgaaaagctcaaggcccttgcaagtgcttcagtgctctcc 240  
Db 599 gtgtcttccttcggtggaagagtgaaaagctcaaggcccttgcaagtgcttcagtgctctcc 658

Qy 241 gctgttcttggtattacagagcgtcgatcttcgtgtgaaaccttcgcctgcctggcg 300  
Db 659 gctgttcttggtattacagagcgtcgatcttcgtgtgaaaccttcgcctgcctggcg 718

Qy 301 tcttcacatctgcgttacccagcatctcgttgccgtttgattgcaactctttaataatc 360  
Db 719 tcttcacatctgcgttacccagcatctcgttgccgtttgattgcaactctttaataatc 778

Qy 361 aaggcagttgcttggcgctcaggtcttcttgggtgtgttcttattgatctccagat 420  
Db 779 aaggcagttgcttggcgctcaggtcttcttgggtgtgttcttattgatctccagat 838

Qy 421 atggtcattgttcttggtgtgagttgttacctttcttcatttcgatttcggtgcagcatt 480  
Db 839 atggtcattgttcttggtgtgagttgttacctttcttcatttcgatttcggtgcagcatt 898

Qy 481 gcttatgccttactgtgttcgcgcacgcagcattgatccagatgcaaccgtgct 540  
Db 899 gcttatgccttactgtgttcgcgcacgcagcattgatccagatgcaaccgtgct 958

Qy 541 ccagtgcttcgaggaacacccaagcgaagcagacccgcagaaatttcaaacgat 600  
Db 959 ccagtgcttcgaggaacacccaagcgaagcagacccgcagaaatttcaaacgat 1018

Qy 601 tccacacatccaggaacaccccttgaccggtgaagctattgcaactgagcagcgcagat 660  
Db 1019 tccacacatccaggaacaccccttgaccggtgaagctattgcaactgagcagcgcagat 1078

Qy 661 gccatgttgcagcgaagccttgctcggcgttgccatgctcccaaccaaggggcag 720  
Db 1079 gccatgttgcagcgaagccttgctcggcgttgccatgctcccaaccaaggggcag 1138

Qy 721 ttagtcttcctggtgagtggaagattgtgtggcatttcccatctggccatcttccga 780  
Db 1139 ttagtcttcctggtgagtggaagattgtgtggcatttcccatctggccatcttccga 1198

Qy 781 gttcgacaaaggctgagtaggtgttccaatgtggatatcttgatgcacattggttccgac 840  
Db 1199 gttcgacaaaggctgagtaggtgttccaatgtggatatcttgatgcacattggttccgac 1258

Qy 841 acagtaaacctcaacgcgcacacacttcaaccgcgtgaagaagcagggcgatgaagtcaaa 900  
Db 1259 acagtaaacctcaacgcgcacacacttcaaccgcgtgaagaagcagggcgatgaagtcaaa 1318

Qy 901 gcaggggagctgctgttgaaatcgatattgatgccattaaaggctgcaggttataggta 960  
Db 1319 gcaggggagctgctgttgaaatcgatattgatgccattaaaggctgcaggttataggta 1378

Qy 961 accagcgcgattgttcttgcgaattacaagaaaacccgacctgtaaaacacttacggtttg 1020  
Db 1379 accagcgcgattgttcttgcgaattacaagaaaacccgacctgtaaaacacttacggtttg 1438

Qy 1021 ggcgaattgaacgggagcccaacctgtcaacgtcgcaagaaagaaagcgggtgccagca 1080  
Db 1439 ggcgaattgaacgggagcccaacctgtcaacgtcgcaagaaagaaagcgggtgccagca 1498

Qy 1081 acaccataagtgaacaccttgagtgctg 1109  
Db 1499 acaccataagtgaacaccttgagtgctg 1527

RESULT 3  
AAH68533/C  
ID AAH68533 standard; DNA; 349980 BP.  
XX  
AC AAH68533;  
XX  
DT 26-SEP-2001 (first entry)  
XX  
C glutamicum coding sequence fragment SEQ ID NO: 7068.  
DE  
XX  
KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;  
KW organic acid synthesis; ds.  
XX  
OS Corynebacterium glutamicum.  
XX  
PN EP1108790-A2.  
XX  
PD 20-JUN-2001.  
XX  
PF 18-DEC-2000; 2000EP-0127688.  
XX  
PR 16-DEC-1999; 99JP-0377484.  
PR 07-APR-2000; 2000JP-0159162.  
PR 03-AUG-2000; 2000JP-0280988.  
XX  
XX (KYOW ) KYOWA HAKKO KOGYO KK.  
XX  
PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;  
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;  
XX  
DR WPI; 2001-376931/40.  
XX  
PT Novel polynucleotides derived from Coryneform bacteria, for identifying  
PT mutation point of a gene, measuring expression of a gene, analysing  
PT expression profile or pattern of a gene and identifying homologous gene  
PT  
PS Disclosure; SEQ ID NO: 7068; 246pp + Sequence Listing; English.  
XX  
XX The present invention provides a number of nucleotide and protein  
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These  
CC are useful for identifying the mutation point of a gene derived from a  
CC mutant of coryneform bacterium, measuring expression amount and  
CC analysing the expression profile or expression pattern of a gene derived  
CC from Coryneform bacterium, and identifying a homologue of a gene derived  
CC from coryneform bacterium. Coryneform bacteria are useful for producing  
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,  
CC particularly L-lysine. The present sequence is a nucleic acid described  
CC in the exemplification of the invention.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from the  
CC European Patent Office.  
XX  
SQ Sequence 349980 BP; 81250 A; 97718 C; 90621 G; 80391 T; 0 other;

Query Match 99.9%; Score 1107.4; DB 22; Length 349980;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1108; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 tatgatttcggtccagtcggtctgctctcggtctgctgtctgtctactcaccatcgctc 60  
Db 110909 TATGATTTCGGTGGTCCAGTCGGCGGTCTGCTCTCGGTCTGGTCTACTCACCACATCGTC 110850

QY 61 atcaactggtgtgcaaccagctctcccccgaattgagctgagctgtttaaaccagggtgga 120  
Db 110849 ATCACTGGTGTGACCAAGTCTCTCCGCCCAATTGAGCTGAGCTGTTTAAACCAAGGTTGGA 110790  
QY 121 tcttcaatcttcgcaacgagcatctatgcttaataatcagccagggtgcgcatgtttggca 180  
Db 110789 TCCTTCACTCTCGCAACGGCATCTATGGCTAAATATCCGCCAGGTCGGCATGTTTGGCA 110730  
QY 181 gttgttctctcgcgaagagtgaaagctcaaggcccttcaggtgcttcaggtgtctcc 240  
Db 110729 GTGTTCTCTCGGGAAGAGTGAAAGCTCAAGGCCCTTCAGGTGCTTCAGGTGTCTCC 110670  
QY 241 gctgttcttggattacggagcctgcgattctcgggtgtaaccttcctgcctgcgctggccg 300  
Db 110669 GCTGTTCTTGGTATTACGGAGCCTGCATCTCGGTGTGAACCTTCGCCCTGCGCTGCGCG 110610  
QY 301 tcttcaatcggtatcggttaacgagcatctaggtggcgctttgattgcaactctttaatatc 360  
Db 110609 TTCTTCACTCGGTATCGGTACCGAGCTATCGGTGGCGCTTTGATTGCACTCTTTAATATC 110550  
QY 361 aaggcagttgctgtggcgctgcaggtttcttgggtgtgtttcttattgattgctccagat 420  
Db 110549 AAGCAGTTCGGTGTGGCGCTGCAGGTTCTTGGGTGTGTTCTATTGATGCTCCAGAT 110490  
QY 421 atggtcatgttcttgggtgtgcagttgttaccttcttcacgttcgcatcgcgcgcagcatt 480  
Db 110489 ATGTCATGTTCTTGGTGTGTCAGTGTGTTACCTTCTTCATCGCATTCGGCGCAGCGATT 110430  
QY 481 gcttatggcctttacitgtgttcgcgaacgagcagcatgattcagatgcaacgctgct 540  
Db 110429 GCTTATGGCCTTTACTTGGTTCCGCCACACGGCAGCATTTGATCCAGATGCAACCGTGTCT 110370  
QY 541 ccagtcctgcaggaacgacaaagcgaagcagaagcagccagcaaatcttcaaacgat 600  
Db 110369 CCAAGTGCCTGCAGACGACCAAGCGAAGCAGAGACACCCGAGAAATTTCAACAGAT 110310  
QY 601 tccaccatcatccagcagcactttgacgggtgaagctattgtcactgagcagcgtcagcgt 660  
Db 110309 TCCACCATCATCCAGGCACCTTTGACCGGTGAAGCTATTGCACTGAGCAGCGTCAGCGAT 110250  
QY 661 gccatgttcgcaaggaagcttgctcggcggtgcgaatcgtcccaacaaaggggcag 720  
Db 110249 GCCATGTTTCCCAAGCGGAAGCTTGGCTCGGGCGTGGCCATCTGCCAACCAAGGGGCGAG 110190  
QY 721 ttagtcttcctggtgagtggaagattgtgtggcattcccatctgccaatgctttcgca 780  
Db 110189 TTAGTTCCTCCGTGAGTGAAGATTTGTTGGCATTCCTCATCTGCCATGCTTTCCGA 110130  
QY 781 gttcgcaccaagcgtgagatggttcccaatgtggatatcttgatgcacattggtttcgac 840  
Db 110129 GTTCGCACCAAGGCTGAGGATGTTCCAAATGTGGATATCTTGATGCACATTTGTTTCGAC 110070  
QY 841 acagtaaacctcaacgagcagcaactttaaccgcgtgaagaagcagggcgatgaagcaaa 900  
Db 110069 ACAGTAAACCTCAACGGCAGCGACTTTAAACCCGCTGAAGAAGCAGGGCGGATGAAGTCAA 110010  
QY 901 gcaagggtgctgtgtgaattcgtattgtgcattaaaggctcaggttatgaggtta 960  
Db 110009 GCAGGGGAGCTGCTGTGTAATTCGATATTGATGCCATTAAGGCTCAGGTTATGAGTTA 109950  
QY 961 accacgcgattgtgttttcgaattacaagaaacacggacacctgttaaacacttacggtttg 1020  
Db 109949 ACCACGCCGATTGTTTTCGAATTAACAAGAAACGGACCTGTAACACTTACGGTTTG 109890  
QY 1021 ggcgaattgaacgagagcgaacctgtcaacgtgcgaagaagaagcgggtgcagca 1080  
Db 109889 GCGGAATTTGAACGGGGAGCGCAACCTGCTCAACGTCGCAAAAGAAAGCGGTGCCAGCA 109830  
QY 1081 acaccataagttgaacaccttgagtgttcg 1109  
Db 109829 ACACCATAAGTTGAACCTTGTAGTGTTCG 109801

RESULT 4  
AAH67869  
ID AAH67869 standard; DNA; 1983 BP.  
XX  
AC AAH67869;  
XX  
DT 26-SEP-2001 (first entry)  
XX  
DE C glutamicum coding sequence fragment SEQ ID NO: 2904.  
XX  
KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;  
KW organic acid synthesis; ds.  
XX  
OS Corynebacterium glutamicum.  
XX  
PN EP1108790-A2.  
XX  
PD 20-JUN-2001.  
XX  
PF 18-DEC-2000; 2000EP-0127688.  
XX  
PR 16-DEC-1999; 99JP-0377484.  
PR 07-APR-2000; 2000JP-0159162.  
PR 03-AUG-2000; 2000JP-0280988.  
XX  
PA (KYOW ) KYOWA HAKKO KOGYO KK.  
XX  
PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;  
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;  
XX  
DR WPI; 2001-376931/40.  
DR P-PSDB; AAG92650.  
XX  
PT Novel polynucleotides derived from Coryneform bacteria, for identifying  
PT mutation point of a gene, measuring expression of a gene, analysing  
PT expression profile or pattern of a gene and identifying homologous gene  
XX  
PS Claim 8; SEQ ID NO: 2904; 246pp + Sequence Listing; English.  
XX  
CC The present invention provides a number of nucleotide and protein  
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These  
CC are useful for identifying the mutation point of a gene derived from a  
CC mutant of coryneform bacterium, measuring expression amount and  
CC analysing the expression profile or expression pattern of a gene derived  
CC from Coryneform bacterium, and identifying a homologue of a gene derived  
CC from Coryneform bacterium. Coryneform bacteria are useful for producing  
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,  
CC particularly L-lysine. The present sequence is a nucleic acid described  
CC in the exemplification of the invention.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from the  
CC European Patent Office.  
XX  
SQ Sequence 1983 BP; 410 A; 508 C; 558 G; 507 T; 0 other;

Query Match 97.8%; Score 1084.4; DB 22; Length 1983;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1085; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 tatgatttcggcgtccagtcgggtctgctcttcggtctggttctactcaccaatcgtc 60  
Db 898 tatgatttcgggtgccagtcggcgtctgctcttcggtctggttctactcaccaatcgtc 957  
QY 61 atcaactggtgtgcaaccagctctcccgccaattgagctgagctgtttaaaccagggtgga 120  
Db 958 atcaactggtgtgcaaccagctctcccgccaattgagctgagctgtttaaaccagggtgga 1017  
QY 121 tcttcatcttcgcaacgagcatctatgcttaataatcgcaggggtgcgcatgtttggca 180  
Db 1018 tcttcatcttcgcaacgagcatctatgcttaataatcgcaggggtgcgcatgtttggca 1077

```
QY 181 gtgttcttctggtgagagtgataaaagctcaaggccttgacagtgcttcaagtgcttc 240
Db 1078 gtgttcttctggtgagagtgataaaagctcaaggccttgacagtgcttcaagtgcttc 1137
QY 241 gctgttcttgggtattacagagctgcgactcttcggtgtgaaccttcgcctgcgctggccg 300
Db 1138 gctgttcttgggtattacagagctgcgactcttcggtgtgaaccttcgcctgcgctggccg 1197
QY 301 ttcttcaatcgggtatcgggtaccgagctatcgggtggcgcttgattgcaactctttaatc 360
Db 1198 ttcttcaatcgggtatcgggtaccgagctatcgggtggcgcttgattgcaactctttaatc 1257
QY 361 aaggcagttgcgttggcgctgcaggtttcttgggttcttgggttcttattgattgctccagat 420
Db 1258 aaggcagttgcgttggcgctgcaggtttcttgggttcttgggttcttattgattgctccagat 1317
QY 421 atggtcatgttcttgggtggtgcagttgttaccttcttcatcgtcattcggcgagcatt 480
Db 1318 atggtcatgttcttgggtggtgcagttgttaccttcttcatcgtcattcggcgagcatt 1377
QY 481 gttattggcctttacttgggttcgcgcgaacggcagcattgatccagatgcaaccgtgtct 540
Db 1378 gttattggcctttacttgggttcgcgcgaacggcagcattgatccagatgcaaccgtgtct 1437
QY 541 ccagtgcttcgaggaacgaccaaagccgaagcagagaagcaccgcagaaatttcaaacgat 600
Db 1438 ccagtgcttcgaggaacgaccaaagccgaagcagagaagcaccgcagaaatttcaaacgat 1497
QY 601 tcacatcatccagcagcactttgaccggtgaagctattgcactgagcagcgtcagcagat 660
Db 1498 tcacatcatccagcagcactttgaccggtgaagctattgcactgagcagcgtcagcagat 1557
QY 661 gcatgtttgcagcggaagcttgctgcggtgcggttgcctatgctcccaaccgaagggcgag 720
Db 1558 gcatgtttgcagcggaagcttgctgcggtgcggttgcctatgctcccaaccgaagggcgag 1617
QY 721 tiagtttctccggtgagtggaagattgtgtggtcattccattccatctgcccattgttcgca 780
Db 1618 tiagtttctccggtgagtggaagattgtgtggtcattccattccatctgcccattgttcgca 1677
QY 781 gttcgacaaagctgagatggttccaatgttgatatcttggatcattgatgcacattggtttcgac 840
Db 1678 gttcgacaaagctgagatggttccaatgttgatatcttggatcattgatgcacattggtttcgac 1737
QY 841 acagtaaaactcaacggcagcgaactttaaccgcgtgaagaagcaggcggatgaagtcaaa 900
Db 1738 acagtaaaactcaacggcagcgaactttaaccgcgtgaagaagcaggcggatgaagtcaaa 1797
QY 901 gcaggggagctgctgtgtaattcgtatattgattgccattaaagctgcagggttatgagta 960
Db 1798 gcaggggagctgctgtgtaattcgtatattgattgccattaaagctgcagggttatgagta 1857
QY 961 accacgccgattgtgttttcgaattacaagaaacggcctgtaaacacattcaggtttg 1020
Db 1858 accacgccgattgtgttttcgaattacaagaaacggcctgtaaacacattcaggtttg 1917
QY 1021 ggcgaataatgaagcgggagccacactgtctcaacgtctgcgaagaaagaaagcgggtgccagca 1080
Db 1918 ggcgaataatgaagcgggagccacactgtctcaacgtctgcgaagaaagaaagcgggtgccagca 1977
QY 1081 acacca 1086
Db 1978 acacca 1983
```

RESULT 5

AAF32543

ID AAF32543 standard; DNA; 5969 BP.

XX

AC AAF32543;

XX

DT 20-APR-2001 (first entry)

XX

```
DE Brevibacterium lactofermentum sucrose PTS enzyme II DNA SEQ ID NO:1.
XX
KW Brevibacterium lactofermentum; sucrose PTS enzyme II; sucrose;
KW phosphoenolpyruvate:carbohydrate phosphotransferase system; glucose;
KW coryneform bacterium; phosphoenolpyruvate-sugar transport system; ds.
XX
OS Brevibacterium lactofermentum.
XX
FH Key Location/Qualifiers
FT CDS 3779..5764
FT /tag= a
FT /product= "sucrose PTS enzyme II"
XX
PN WO200102584-A1.
XX
PD 11-JAN-2001.
XX
PF 30-JUN-2000; 2000WO-JP04348.
XX
PR 02-JUL-1999; 99JP-0189512.
XX
PA (AJIN ) AJINOMOTO CO INC.
XX
PI Izui M, Sugimoto M, Nakamatsu T, Kurahashi O;
XX
XX WPI; 2001-138150/14.
DR P-PSDB; AAB69080.
XX
XX Gene encoding sucrose phosphoenolpyruvate-sugar transport system enzyme
PT II obtained by cassette ligation-mediated amplification of downstream
PT domain of coryneform bacterium sucrose gene, with sucrose-binding
PT activity .
XX
PS Claim 3; Page 22-29; 45pp; Japanese.
XX
CC The present sequence encodes the Brevibacterium lactofermentum sucrose
CC PTS (phosphoenolpyruvate:carbohydrate phosphotransferase system or
CC phosphoenolpyruvate-sugar transport system) enzyme II, which has
CC sucrose-binding activity. A coryneform bacteria produced with the
CC sucrose PTS enzyme II gene can have more efficient sugar uptake, and
CC improved amino-acid and nucleic acid productivity. The sucrose PTS gene
CC and it's disrupted gene, such as one without the sucrose PTS function,
CC can be used to produce new breeds of coryneform bacterial strains to
CC uptake sugar more efficiently e.g. glucose only or and sucrose, and can
CC have improved amino-acid and nucleic acid productivity.
XX
SQ Sequence 5969 BP; 1388 A; 1575 C; 1578 G; 1428 T; 0 other;
```

Query Match 94.7%; Score 1049.8; DB 22; Length 5969;  
Best Local Similarity 96.7%; Pred. No. 8.7e-314;  
Matches 1072; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 1 tatgatttcggcggtccagtcggtctgtctcttctggtctgttctactcaccatcgctc 60  
Db 4676 tatgatttcggcggtccagtcggtctgtctcttctggtctgttctactcaccatcgctc 4735  
QY 61 atcaactggctgcaccagctctcccgccaattgagctggagctgtttaaccagggtgga 120  
Db 4736 atcaactggctgcaccagctctcccgccaattgagctggagctgtttaaccagggtgga 4795  
QY 121 tccttcaatcttcgaacggcgcattatggttaataatgccaccagggtgcggtgtttggca 180  
Db 4796 tccttcaatcttcgaacggcgcattatggttaataatgccaccagggtgcggtgtttggca 4855  
QY 181 gtgttcttctggcgaagagtgataaaagctcaaggccttcagggtgcttcaagtgcttc 240  
Db 4856 gtgttcttctggcgaagagtgataaaagctcaaggccttcagggtgcttcaagtgcttc 4915  
QY 241 gctgttcttgggtattacggagcctgcgactcttcggtgtgaaccttcgcctgcgctggccg 300  
Db 4916 gctgttcttgggtattacggagcctgcgactcttcggtgtgaaccttcgcctgcgctggccg 4975

```
QY 301 ttcttcacgtatcggtaccgcagctatcggtgctgttggcttggatgcactctttaaatac 360
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 4976 ttctacattggtatcggtaccgcagctatcggtgctgttggcttggatgcactcttggatc 5035
QY 361 aaggcaattgctggcgctgcaggtttcttctggtgttggcttattgatgtcccgat 420
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 5036 aaggcagttgctggcgctgcaggtttcttctggtgttggcttattgatgtcccgat 5095
QY 421 atagtcattgttctggtgtgcaggtgttaccttcttcattcgcattcggcgagcgatt 480
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 5096 atggcattgttctggttctggtgagttaccttcttcattcgcattcggcgagcgatt 5155
QY 481 gcttatggcttacttggctgcgcgaacggcgagctattgatccagatgcaacgcgtgct 540
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 5156 gcttatggcttacttggctgcgcgaacggcgagctattgatccagatgcaacgcgtgct 5215
QY 541 ccagtgctgcaggaagacaaagccgaagcgaagcgaagcgaagcgaagcgaagcgaagcga 600
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 5216 ccagtgctgcaggaagacaaagccgaagcgaagcgaagcgaagcgaagcgaagcgaagcga 5275
QY 601 tccaccatcatccagcagcttggacgggtgaaagctattgacatgagcagcgtcagcgat 660
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 5276 tccaccatcatccagcagcttggacgggtgaaagctattgacatgagcagcgtcagcgat 5335
QY 661 gccatgtttgccagcgaaagcttggctcgggcttggccttccatcgtcccaaccaaaggcgag 720
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 5336 gccatgtttgccagcgaaagcttggctcaggtgttgcgctcctccaccaaaggcgag 5395
QY 721 tgaattctcgggtgagtgaaagattggtggcgatcccatctgccatctgcccattgcttcgca 780
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 5396 ctggttccaccagtgagcggaaagatggtggccttcccatctggtcagcgttctcgca 5455
QY 781 gttccgaccaaagctgaggtgttccaatgtgatatcttgatctgacattggtttcgac 840
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 5456 gtccgactaaggctggaggtgttccaatgtgatatcttgatgacattggtttcgac 5515
QY 841 acagtaaacctcaacggcagcactttaaccgctgaagcagcggcgatgaagtcaaa 900
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 5516 accgtaaacctcaacggcagcactttaaccgctgaagcagcggcgatgaagtcaaa 5575
QY 901 gcaggggagctgctgtgtaattgatgatgcattgacccattgaaggtcgaggttatgaggt 960
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 5576 gcaggggagctgctgtgtaattgatgatgcattgacccattgaaggtcgaggttatgaggt 5635
QY 961 accacgctgattgtgttgcgaattacaagaacacggacctgttaaacacttacggtttg 1020
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 5636 accacgctgattgtgttgcgaattacaagaacacggacctgttaaacacttacggtttg 5695
QY 1021 ggcgaattgaaggggagccacctgtcaacgctcgaagaagcgaagcgggtgccagca 1080
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 5696 ggcgaattgaaggggagccacctgtcaacgctcgaagaagcgaagcgggtgccagca 5755
QY 1081 acaccataagtgaacacctgagtgctcg 1109
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 5756 acaccataagtgaacacctgagtgctcg 5784
```

## RESULT 6

AA12968

ID AA12968 standard; DNA; 5840 BP.

XX AC

XX AC

XX AC

XX AC

DT 19-MAR-1999 (first entry)

XX DE Enterococcus faecalis genome contig SEQ ID NO:31.

XX KW Enterococcus faecalis; contig; detection: Enterococcal infection;

XX KW vaccine; attenuation; computer readable medium; ds.

XX OS Enterococcus faecalis.

XX OS

XX PN

PN WO9850555-A2.

```
XX 12-NOV-1998.
PD ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX 04-MAY-1998; 98WO-US08985.
PF ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX 14-NOV-1997; 97US-0066009.
PR ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
PR 06-MAY-1997; 97US-0044031.
PR 16-MAY-1997; 97US-0046655.
XX (HUMA-) HUMAN GENOME SCI INC.
PA ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX Barash SC, Dillon PJ, Kunsch CA;
PI WPI; 1999-045171/04.
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX New isolated Enterococcus faecalis polynucleotides and polypeptides
PT - used to develop products for the detection of Enterococcus and for
PT use in vaccines for prevention or attenuation of Enterococcus
PT infection.
XX Claim 1; Page 389-392; 2084pp; English.
PS A computer readable medium has been developed which has recorded on it
XX 982 nucleotide sequences isolated from the Enterococcus faecalis genome.
CC AAX12938 to AAX13919 represent these nucleotide sequences which are
CC primary nucleotide sequences, also known as contigs. The computer-based
CC system can identify fragments of the Enterococcus faecalis genome with
CC commercial importance. The products can be used to detect the presence
CC of Enterococcus faecalis in samples. They can also be used for
CC diagnosing Enterococcal infection in an animal and monitoring
CC progression of disease, and for identifying agents which can be used to
CC modulate the growth or pathogenicity of Enterococcus faecalis, or
CC another related organism, in vivo or in vitro. In particular the
CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences
CC can be used in vaccines to prevent or attenuate an Enterococcal
CC infection.
XX Sequence 5840 BP; 1848 A; 944 C; 1256 G; 1779 T; 13 other;
```

Query Match 11.6%; Score 128.6; DB 20; Length 5840;  
Best Local Similarity 48.3%; Pred. No. 5.2e-29;  
Matches 470; Conservative 0; Mismatches 479; Indels 24; Gaps 3;

```
QY 20 tcggcgctgctcttcggtctgcttactcaccaatcgctcatcactctgctgcacacgt 79
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1959 tcgcaggggtattaatgggttcgtgttggaagcttggtaattgttggatgattggg 2018
QY 80 ccttcccgcccaattgagctggagctgtttaaccagggtggatccttcatcttccgacgg 139
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2019 gcttgtaccaattatgatgttaaaccttaacacagggtggcgatcacgatggtaccgatgt 2078
QY 140 catctatggtaataatgcgccagggtgcggcatgtttggcaggtgttcttccctggcgaga 199
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2079 tattaccagcgttattgcaacaggcgggctgttagctgtcttcttcttaacaaaaa 2138
QY 200 gtgaaaagctcaaggccttgcaggcttcacagctgctccgctgttcttcttcttcttcttctt 259
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2139 atgtgaacataaaagggttggcttcttcttcttcttcttcttcttcttcttcttcttcttctt 2198
QY 260 agcctgcgactcttcggtgtgaaccttcgctgcgtgcgtggcgttcttcttcttcttcttcttctt 319
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2199 aaccaactgtatgctgctgactttaccattgaaaaaaccaatttatgcactgtattgt 2258
QY 320 ccgcagctatcggtggcgcttggatgcactcttataatcaagcaggttgcgttggcg 379
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2259 gtggcggtatcggtgtgcatttggctatgaatcactgaaacacactttacgtttggt 2318
QY 380 ctgcaggtttcttctggtgttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 439
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2319 tggtagtattgagcttgcgttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 2378
QY 440 gtgcaggtgttaccttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 499
```

```

Db 2379 tgattactggtgcaatggtgcccgaattgctttatcatgctgttgccttaacggttg 2438
Qy 500 ttgcccgaacgagcagcattgatccagatgcaaccgctgctccagtgctgcaggaacga 559
Db 2439 tcttac-----gtttgaagatcaacctaatccagaaacagcaactgaaaaaacag 2489
Qy 560 ccaagccgaagcagaagcaccgcagcaattttcaaacgattccacatccacaggca- 618
Db 2490 aaactgataagatggtggcaccctgtaaaacgaatcaagaagacaaaatttttagcaa 2549
Qy 619 --cctttgaccggtgaagctattgcactgagcagcgtgcagcgtgccaatggttggcagcg 676
Db 2550 gtcaacttcaagtgaaattttaccgctagaaaaagtaacaagaccctgttttggcttcag 2609
Qy 677 gaaagcttggtcgggcgttgccatcgctcccaacaaaggggcagtgtagttctccggtga 736
Db 2610 gtgctttaggaaaagggtgtgcaattgagccgactgaaggcgaactgtatgcaccgcag 2669
Qy 737 gtgaaagattggtggcattcccatctgccaatgcttgcagttgcagttgcacaaaggctg 796
Db 2670 atgtgaaatcaccacattattccgacaggaacatgctgttggcttgacgacaacagagg 2729
Qy 797 aggatggttccaatgtgatatctgtgacacattggtttcgacacagtaaacctcaacg 856
Db 2730 cggttgat-----tattaatgcattatggcatggatgatacgggtcgaattagatg 2777
Qy 857 gcacgcaacttaaccgctgaaagcagcagggcgatgaagtcaaaagcagggagctgctgt 916
Db 2778 gtaagggcttgaattatcagtgaaacaagggtgattctgttcaaaaaggagattgctag 2837
Qy 917 gtgaattcgattgatgcattaaaggctgcaggttatgaggttaacacgcgcgattgtg 976
Db 2838 ttactttgatctgctgccattaaagaagctggttatccggttagttacacccgattgtg 2897
Qy 977 ttccgaattacaa 989
Db 2898 taacgaatacga 2910

RESULT 7
AAH54100/c
ID AAH54100 standard; DNA; 2913 BP.
AC AAH54100;
XX
XX 03-SEP-2001 (first entry)
DE S. epidermidis genomic polynucleotide sequence SEQ ID NO:3464.
KW Staphylococcus epidermidis SRI strain; infection; diagnosis;
vaccination; endocarditis; ds.
XX Staphylococcus epidermidis.
XX WO200134809-A2.
XX 17-MAY-2001.
XX 09-NOV-2000; 2000WO-US30782.
XX 09-NOV-1999; 99US-0164258.
XX (GLAX ) GLAXO GROUP LTD.
XX Kimmerly WJ;
XX WPI; 2001-316495/33.
XX Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
PT useful for vaccinating against infections, e.g. endocarditis -
XX
XX Claim 8; Page 1030-1031; 2188pp; English.

```

```

XX
CC AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.
CC (I) and (II) can have antibacterial activity and therefore can be used
CC in vaccination. The nucleic acids (I) may be used to produce the
CC S. epidermidis polypeptides (II) via the production of vectors
CC containing them which are used to produce hosts cells which express the
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
CC used to vaccinate subjects and to raise antibodies against the bacteria.
CC The polypeptides may also be used to assay for other inhibitors of their
CC activity and therefore identify compounds that may be used for the
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
CC AAH55090 represent specifically claimed S. epidermidis genomic DNA
CC polynucleotide sequences from the present invention. AAH55091 to
CC AAH55098 represent oligonucleotide sequences and primers which are used
CC in the exemplification of the present invention.
CC N.B. The present invention specifically claims all the polynucleotide
CC sequences given in the sequence listing of the present specification,
CC however the sequence listing only goes up to SEQ ID NO:4454 so even
CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
CC no sequences are present for SEQ ID NO:4455 to 4464.
XX
SQ Sequence 2913 BP; 1072 A; 480 C; 419 G; 942 T; 0 other;

Query Match 11.1%; Score 123.2; DB 22; Length 2913;
Best Local Similarity 59.0%; Pred. No. 1.7e-27;
Matches 256; Conservative 0; Mismatches 163; Indels 15; Gaps 2;

Qy 1 tatgattcgccggtccagtcggctgctcttcggtgctgctggtctactcaccacgctc 60
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2274 TATGAATTGGTGGAGCTATTGGTGGACTTATATTTGGTTTATATATATGCGCCATCGTC 2215
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 61 atcactgctgcacagctccctccgcgaattgagctg-----gagctgctt 108
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2214 ATTACAGGAATGCACCATAGCTTTATTGAGTTGAAACGACATTAATTGCTGATCGGACT 2155
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 109 aaccaggttgatccttcattcgaacggcatctatgctaatatcccccaggtgcg 168
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2154 AAACAGGTGGTTCATTATCTCCCAATFCGCAACGATGTCAATATTGCACAAGGTGT 2095
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 169 gcatgtttggcagtgcttct---tccctggcgaagagtgaaagctcaaggccttcgaggt 225
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2094 GCAGCTTTAGCTGCATCTTTATCATTAAGCAATAATAAATAAAGTGTGTGCTTCC 2035
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 226 gcttcaggtgctccgctgtcttcttggtattacggagcctgcgactctcgggtgaaacctt 285
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2034 GCGGCGGGTATTTCAGCTTTACTAGGAATTACAGAACCAACCAATGTTTGGTGTCAATCTT 1975
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 286 cgcctgcctggcgttcttcctcgtatcggtaccgagctatcggtgagcgttggatt 345
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1974 AAATTGAGATATCCATTATAGGTGCTGTTGAGGATCAGGTATAGTGCAGCTTATATT 1915
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 346 gcactctttaatacaaggcaggttgcgttggcgctgcaggtttcttgggtgtgtttct 405
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1914 TCATTCTTCAAGTAAAGCGATAGCGCTTGTCACAGCTGGATTACCTGGATTATATCT 1855
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 406 attgatgtccaga 419
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1854 ATAAATCCTACACA 1841
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 8
AAV52334/c
ID AAV52334 standard; DNA; 3895 BP.
XX
XX AC AAV52334;
XX
XX 23-OCT-1998 (first entry)
DE Streptococcus pneumoniae genome fragment SEQ ID NO:201.
XX
XX Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;

```





```

PR 07-JUL-1992; 92JP-0179719.
XX (FUSO ) FUSO PHARM IND LTD.
PA (OHNO/) OHNO T.
XX
XX
PI Eda S, Matsuhisa A, Ohno T, Uehara H;
XX WPI; 1994-035086/04.
XX
XX
XX The nucleotide sequence of a 3615 bp probe obtained by digestion of
CC Escherichia coli genomic DNA with the restriction enzyme HindIII.
CC The probes (AAQ55133-48 and AAQ55752-58) represent other probes derived,
CC by HindIII digestion of the genomes of Staphylococcus aureus,
CC S.epidermis, Enterococcus faecalis, Pseudomonas aeruginosa, E.coli,
CC Klebsiella pneumoniae or Enterobacter cloacae. The probes can be used
CC to detect their respective microorganisms in clinical samples.
XX
XX Sequence 3615 BP; 966 A; 962 C; 805 G; 882 T; 0 other;
SQ
Query Match 8.2%; Score 91; DB 15; Length 3615;
Best Local Similarity 53.7%; Pred. No. 1.7e-17;
Matches 240; Conservative 0; Mismatches 195; Indels 12; Gaps 2;
QY 610 attcaggacaccttgaccggtgaagctattgcaactgagcagcgatgcagtggttt 669
DB 1798 ATCTGTTACCGATGACGGAGAGATTGTCGCTCATTCACGTCGCTGATACCGGTTT 1739
QY 670 gccagcggaagcttggtcggcggttgcacatgctccatcgtcccaacgaagggttagttct 729
DB 1738 GCCAGTGGCGCTGTTGGGTAAAGGTAATTGCCAATCTGCCCTCGGTTGGTGAAGTCGCTTCT 1679
QY 730 ccggtgagtggaagattggttggtgcatcccatcgtggtcagtgcttcgagttcgacc 789
DB 1678 CCGTGTGGCGGTGCGAATTGCTGTTGTTGGCCACATTAACA-----CGCCATTGGCAAT 1625
QY 790 aaggctgagtggttcccaatgtggatatcttgatgcacattggtttcgacacagtaaac 849
DB 1624 GAGTCAGATGATGG-----TGTGGAGATCCGTGATTGTCGATCGATCGACACCGTAATA 1571
QY 850 ctcaacggcagcactttaaccgcgtgaagcagcagggcgatgaagtcaaacgagggag 909
DB 1570 CTGGACGGCAAAATTTCTTTTCCGCTCACGTCAACGTGGTGACAAGGTCAATACAGCGCAT 1511
QY 910 ctgctgtgtaattcgatattgattgacattaggtggtggttattgagtaaccacgccc 969
DB 1510 CGGCTGATTTCTTTGATATCCCTGCTATTCGGGAGCGCGGATTTGATCTGACGACGCG 1451
QY 970 attgtgtttcgaattacagaacacccgacactgttaacacacttgcgtttggcgcaaat 1029
DB 1450 GTATTAAATCAGTAATAGCATGATTATTACGGAGCATATTACCCACCGCAGCGCGCAGATA 1391
QY 1030 gaagcgggagcacaactgctcaagtc 1056
DB 1390 AGCGCAGGTGAACCGCTGTATCCATC 1364
RESULT 10
AAV52163/c
ID AAV52163 standard; DNA; 9769 BP.
XX
XX
XX AAV52163;
XX
DT 23-OCT-1998 (first entry)
XX
XX Streptococcus pneumoniae genome fragment SEQ ID NO:30.
XX
Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
computer readable medium; vaccine; pharmaceutical composition; ds.
Streptococcus pneumoniae.
W09818931-A2.
07-MAY-1998.
30-OCT-1997; 97WO-US19588.
31-OCT-1996; 96US-0029960.
(HUMA-) HUMAN GENOME SCI INC.
Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M;
Kunsch CA, Rosen CA;
WPI; 1998-272225/24.
Computer-readable medium with recorded Streptococcus pneumoniae
polynucleotide sequences - useful in diagnostic kits and assays, and
pharmaceutical compositions and vaccines for Streptococcus
pneumoniae
Claim 1; Page 318-324; 1409pp; English.
The present invention describes a computer readable medium which has
the nucleotide sequences SEQ ID NO:1 to 391 (AAV52134 to AAV52524)
recorded on it, or a representative fragment or a sequence at least 95%
identical to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in
SEQ ID NO:1 to 391 (AAV52134 to AAV52524) are genomic fragments from
Streptococcus pneumoniae. The present invention also describes an
isolated nucleic acid molecule encoding a homologue of any of the
fragments of the S.pneumoniae genome (SEQ ID NO:1 to 391) where the
nucleic acid molecule is produced by a process comprising: (a) screening
a genomic DNA library using as a probe a target sequence defined by any
of the sequences in SEQ ID NO:1 to 391, identifying members of the
library which contain sequences that hybridize to the target sequence and
isolating the nucleic acid molecules from the members; or (b) isolating
mRNA, DNA or cDNA produced from an organism, amplifying nucleic acid
molecules whose nucleotide sequence is homologous to amplification
primers derived from the fragment of the S. pneumoniae genome to prime
the amplification and isolating the amplified sequences. The computer
readable medium can be used in a computer-based system for identifying
fragments of the S. pneumoniae genome of commercial importance, or
expression modulating fragments of the S. pneumoniae genome. Products
from the present invention can be used in diagnosis kits and assays, and
pharmaceutical compositions and vaccines for S. pneumoniae.
Sequence 9769 BP; 2918 A; 2141 C; 1792 G; 2916 T; 2 other;
Query Match 8.1%; Score 90.2; DB 19; Length 9769;
Best Local Similarity 46.7%; Pred. No. 5.1e-17;
Matches 490; Conservative 0; Mismatches 523; Indels 36; Gaps 5;
QY 20 toggcggtctgctcttggtctggtctactcacaatcgctcaactggtggtcgcacagt 79
DB 8843 TTGGTGGTTTCTTGATTGGTGGGGTTTCAACAAATTTGATCGTGTGACGGTGTGCACACA 8784
QY 80 ccttcccgccaattgagctgga--gctgtttaaccagggtggtatccttcatctcgcaa 136
DB 8783 TCTTCAACTTGGTTGAGTGCAATTTACTTCTGCTGACCAATGTAACCCATTCAACGCTA 8724
QY 137 cggcatctatggttaataatgcccgaggtgcggcagtggttggcagtggttcttctcgcca 196
DB 8723 TCATCACAGCTGCTATCACAGCTCAAGGTCTGCTACTACTGTTGCGTGTGTTAAACAA 8664
QY 197 agagtgaagagctcaaggcgcttcaggtgcttcaggtgcttcgctgtcttggtatta 256
DB 8663 AAAATCCAAAACCTGAAACACACTTGTCTTCCCGGTGCTCTTCTGCTTCTCTCTAGGTATTA 8604

```



```

RESULT 12
AAV52296
ID AAV52296 standard; DNA; 8494 BP.
XX
AC AAV52296;
XX
DT 23-OCT-1998 (first entry)
XX
DE Streptococcus pneumoniae genome fragment SEQ ID NO:163.
XX
KW Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
computer readable medium; vaccine; pharmaceutical composition; ds.
XX
OS Streptococcus pneumoniae.
XX
PN WO9818931-A2.
XX
PD 07-MAY-1998.
XX
PF 30-OCT-1997; 97WO-US19588.
XX
PR 31-OCT-1996; 96US-0029960.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M;
PI Kunsch CA, Rosen CA;
XX
DR WPI; 1998-272225/24.
XX
Computer-readable medium with recorded Streptococcus pneumoniae
polynucleotide sequences - useful in diagnostic kits and assays, and
pharmaceutical compositions and vaccines for Streptococcus
pneumoniae
Claim 1: Page 1052-1057; 1409pp; English.
XX
The present invention describes a computer readable medium which has
the nucleotide sequences SEQ ID NO:1 to 391 (AAV52134 to AAV52524)
recorded on it, or a representative fragment or a sequence at least 95%
identical to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in
SEQ ID NO:1 to 391 (AAV52134 to AAV52524) are genomic fragments from
Streptococcus pneumoniae. The present invention also describes an
isolated nucleic acid molecule encoding a homologue of any of the
fragments of the S. pneumoniae genome (SEQ ID NO:1 to 391) where the
nucleic acid molecule is produced by a process comprising: (a) screening
a genomic DNA library using as a probe a target sequence defined by any
of the sequences in SEQ ID NO:1 to 391, identifying members of the
library which contain sequences that hybridise to the target sequence and
isolating the nucleic acid molecules from the members; or (b) isolating
mRNA, DNA or cDNA produced from an organism, amplifying nucleic acid
molecules whose nucleotide sequence is homologous to amplification
primers derived from the fragment of the S. pneumoniae genome to prime
the amplification and isolating the amplified sequences. The computer
readable medium can be used in a computer-based system for identifying
fragments of the S. pneumoniae genome of commercial importance, or
expression modulating fragments of the S. pneumoniae genome. Products
from the present invention can be used in diagnosis kits and assays, and
pharmaceutical compositions and vaccines for S. pneumoniae.
XX
Sequence 8494 BP; 2532 A; 1570 C; 1870 G; 2522 T; 0 other;

Query Match 6.0%; Score 67; DB 19; Length 8494;
Best Local Similarity 47.7%; Pred. No. 7e-10;
Matches 239; Conservative 0; Mismatches 250; Indels 12; Gaps 1;

Qy 507 caacggcgagcattgaccagatgaacgcgtgctccagtcgaggaacacccaagc 566
Db 1609 ccaagatattcttgatcagtggaattcttctgaaactcttccaagccaatgactga 1668
Qy 567 gaaagcagaagcaccgcagaaattttcaaacgattccaccattccagggcactttgac 626

```

Db 1669 agcacacacaaacactgttcacttccaaagatcttactgaggaagtttactcagtagcaga 1728

Qy 627 cggatgaagctattgcactgagcagcgtcagcagatgccatgtttgcccagcggaaaagcttgg 686

Db 1729 cggatgaagctattgtcttggaaacagtaaaagatccagtagtttgcataaaatgatggg 1788

Qy 687 ctccggcgttcacatcgtcccaacacaaagggcaggttagtttctccggtgagtggaagat 746

Db 1789 tgatggatttgtagtagaaccttgcaaatggaacacatgctatctccagtttcaggtactgt 1848

Qy 747 tgtgtggcattcccatctggtccatcttgcagcttcgcacccacaggtcaggtggttc 806

Db 1849 gtcaagcattctcccaacaaacatgctttgtattgtgacggaagcagg----- 1899

Qy 807 caatgtggatatcttgatgcacattggtttcacacagtaaacctcaacggcagcactt 866

Db 1900 ---tcttgaagtattggttcacattggttgacacagtaagcttgaaggtaaacatt 1956

Qy 867 taaccgcgtgaagaagcagggcgatgaagtaacagcagggagctgctgtgtgaattcga 926

Db 1957 tacagttcatgttctgaaggacaaaagtgtagcagcagagatctctctgtcacagctga 2016

Qy 927 tatgatgccattaaagctgcaggttatgaggttaaccacgcgcgattgttttcgaattca 986

Db 2017 ctgtgagctatccgtgcagcagcagctgaaaacttcaacagtagttgttctcacaaatgg 2076

Qy 987 caagaaaacccggacctgtataa 1007

Db 2077 tgatgcaattaaatcagtttaa 2097

RESULT 13

AAH81338

ID AAH81338 standard; DNA; 1947 BP.

XX

AC AAH81338;

XX

DT 21-SEP-2001 (first entry)

XX

DE Escherichia coli protein encoding nucleotide sequence SEQ ID NO:137.

XX

KW Escherichia coli; identification; proliferation; microorganism;

KW antimicrobial; antibacterial; antibiotic; gene therapy; diagnosis;

KW bacterial growth inhibition; ds.

XX

OS Escherichia coli.

XX

PN WO200148209-A2.

XX

PD 05-JUL-2001.

XX

PF 19-DEC-2000; 2000WO-US34419.

XX

PR 23-DEC-1999; 99US-0173005.

XX

PA (ELIT-) ELITRA PHARM INC.

XX

PI Forsyth RA, Ohlsen KL, Zyskind JW;

XX

DR WPI; 2001-457376/49.

DR P-PSDB; AAG98282.

XX

PT Novel nucleic acids encoding proteins required for Escherichia coli

PT proliferation, useful for screening for antimicrobial agents -

XX

PS Claim 9; Page 180-183; 596pp; English.

XX

CC The present invention describes a purified or isolated nucleic acid

CC sequence (I) consisting essentially of one of the 93 nucleotide sequences

CC given in AAH81202 to AAH81294, where expression of the nucleic acid in a

CC microorganism is capable of inhibiting proliferation of a microorganism.

CC (I) have antibacterial and antibiotic activities, and can be used in

CC gene therapy. Expression of (I) in a microorganism inhibits proliferation





**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: March 22, 2002, 08:49:30 ; Search time 2009.21 Seconds  
(without alignments)  
9105.754 Million cell updates/sec

Title: US-09-604-231-3  
Perfect score: 1109  
Sequence: 1 tatgatttcggcggtccagt.....gttgaaccttgagtgttcg 1109

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

- GenEmbl.\*  
1: gb.ba.\*  
2: gb.htg.\*  
3: gb.in.\*  
4: gb.om.\*  
5: gb.ov.\*  
6: gb.pat.\*  
7: gb.ph.\*  
8: gb.pl.\*  
9: gb.pr.\*  
10: gb.ro.\*  
11: gb.sts.\*  
12: gb.sy.\*  
13: gb.un.\*  
14: gb.vi.\*  
15: em.ba.\*  
16: em.fun.\*  
17: em.hum.\*  
18: em.in.\*  
19: em.om.\*  
20: em.or.\*  
21: em.ov.\*  
22: em.pat.\*  
23: em.ph.\*  
24: em.pl.\*  
25: em.ro.\*  
26: em.sts.\*  
27: em.sy.\*  
28: em.un.\*  
29: em.vi.\*  
30: em.htgo\_hum.\*  
31: em.htgo\_inv.\*  
32: em.htgo\_rod.\*  
33: em.htg\_hum.\*  
34: em.htg\_inv.\*  
35: em.htg\_rod.\*  
36: em.htg\_Other.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

8

Result No.	Score	Query Match	Length	DB	ID	Description
1	1109	100.0	1109	6	AX069136	AX069136 Sequence
2	1107.4	99.9	1527	6	AX069134	AX069134 Sequence
c	1107.4	99.9	349980	6	AX127152	AX127152 Sequence
3	1084.4	97.8	1983	6	AX122988	AX122988 Sequence
5	165.6	14.9	2508	1	STRSCRA	M2711 Streptococ
c	138.4	12.5	21838	1	PDCRAFOPER	L32093 Pedococcus
c	138.2	12.5	21838	1	PPSURFOP	Z32771 P.pentose
8	135.2	12.5	296950	1	AP001508	AP001508 Bacillus
9	135.2	12.2	2655	1	SXSCRA	X69800 S.xylosus s
c	133	12.0	5800	1	LLZ97015	Z97015 Lactococcus
c	127	11.5	10264	1	AE004395	AE004395 Vibrio ch
c	126.6	11.4	10085	1	AE006222	AE006222 Pasteurel
c	123.2	11.1	2913	1	AF269422	AF269422 Staphyloc
c	123.2	11.1	2913	6	AX144742	AX144742 Sequence
c	121	10.9	341350	1	AP003365	AP003365 Staphyloc
c	120.4	10.9	151947	2	SPNEU1902	AL449934 Streptoco
17	120	10.8	2955	1	VBSCRAK	M76768 Vibrio algi
18	119.4	10.8	4158	1	BSTREAPR	Z54245 B.subtilis
c	118.8	10.7	14515	1	AE007479	AE007479 Streptoco
c	115.8	10.4	22197	1	D83967	D83967 Bacillus su
c	115.8	10.4	37900	1	D86417	D86417 Bacillus su
22	115.8	10.4	208430	1	BSUB0005	Z99108 Bacillus su
c	110.4	10.0	299850	1	AP001514	AP001514 Bacillus
24	104.8	9.4	3518	1	BACSACP	J03006 B.subtilis
25	104.8	9.4	97015	1	BSENGR	X73124 B.subtilis
c	104.8	9.4	212150	1	BSUB0020	Z99123 Bacillus su
27	104.4	9.4	4487	1	AF229829	AF229829 Pseudomon
c	101.4	9.1	5065	1	ERWBGPA	M81772 Erwinia chr
c	100.6	9.1	2819	1	BACISPO	D37921 Alkalophthol
30	98.4	8.9	8006	1	AF206272	AF206272 Streptoco
31	97.8	8.8	2543	1	BSTREPA	X80203 B.subtilis
c	97.8	8.8	292550	1	AP001513	AP001513 Bacillus
c	92.8	8.4	12086	1	AE006630	AE006630 Streptoco
c	92.2	8.3	136254	1	ECOW82	L10328 E. coli; th
35	91	8.2	2134	1	ECOBGLC	M15746 E.coli bgIC
c	91	8.2	3615	6	AR012068	AR012068 Sequence
c	91	8.2	3615	6	AR025193	AR025193 Sequence
c	91	8.2	3615	6	AR038650	AR038650 Sequence
c	91	8.2	3615	6	AR068497	AR068497 Sequence
c	91	8.2	5270	1	ECOBGLQ	M16487 E.coli bgl
c	91	8.2	12989	1	AE000449	AE000449 Escherich
c	90.2	8.1	10371	1	AE007465	AE007465 Streptoco
43	88.6	8.0	2343	3	AF309637	AF309637 Leishmani
44	86.6	7.8	300750	1	AP001509	AP001509 Bacillus
45	85.8	7.7	4621	1	KPSCRYAB	X57401 K.pneumonia

ALIGNMENTS

RESULT	1
AX069136	
LOCUS	AX069136 1109 bp DNA
DEFINITION	Sequence 3 from Patent WO0102583.
ACCESSION	AX069136
VERSION	AX069136.1
KEYWORDS	GI:12579018
SOURCE	Corynebacterium glutamicum.
ORGANISM	Corynebacterium glutamicum. Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacterineae; Corynebacteriaceae; Corynebacterium.

REFERENCE	1 (bases 1 to 1109)
AUTHORS	Pompeius M., Kroeger, B., Schroeder, H., Zelder, O. and Haberhauer, G.
TITLE	Orynebacterium glutamicum genes encoding phosphoenolpyruvat e: sugar phosphotransferase system proteins
JOURNAL	Patent: WO 0102583-A 3 11-JAN-2001;
FEATURES	BASF AKTIENGESellschaft (DE)
SOURCE	Location/Qualifiers 1. .1109





|||||  
Db 419 TATGATTTCCGGTGGTCCAGTCGGCGTCTGCTCTCGGCTCGTCTACCTACCAATCGTC 478  
QY 61 atcaatggtctgcacacgctccctccgcgaattgagctggagctggttttaaccagggtgga 120  
|||||  
Db 479 ATCACTGGTCTGCACCAAGTCTCCCGCAATTAGCTGAGCTGTTAAACAGGGTGGA 538  
QY 121 tcttcacatctgcacacgcatctatggttaataatgcaccagggtgcgcgcatgtttggca 180  
|||||  
Db 539 TCTTCATCTTCGCAACGGCACTATPFGGTAATATGCCCGCAGGGTGGCATGTTTGGCA 598  
QY 181 gtgttctctgcggaagagtgtaaaagctcaaggcccttgacaggtgcttcagggtgctcc 240  
|||||  
Db 599 GTGTTCTTCTGCGGAAGAGTGAAGAGCTCAAGGGCTTGCAGGTGCTTCAGGTGCTCC 658  
QY 241 gctgttcttggtattacagagcctgcgatctctcggtgtgaaacctgcctgcctggccg 300  
|||||  
Db 659 GCTGTGTTCTTGGTATTACGGAGCGTCCGATCTCGGTGTGAACCTTGCCTGGCGTGGCG 718  
QY 301 tcttcacatcggtatcggtaccagcgtatcggtgagcgtttgattgacatctttaataac 360  
|||||  
Db 719 TCTTCATCGGTATCGTACCAGCATATCGGTGGCGCTTGTATGCACTCTTTAATATC 778  
QY 361 aaggcagttgctggtggcgctgcaggtttcttggtgtgtttctattgatgctccagat 420  
|||||  
Db 779 AAGGCAGTTGCGTTGGCGCTGCAGGTTCTTGGGTGTTGTTCTATTATGCTCCAGAT 838  
QY 421 atggtcatgttcttggtgtggtcagttgttaccttcttcacatgcctggcgcagcatt 480  
|||||  
Db 839 ATGGTCATGTTCTGTTGTGTGTCAGTTGTTAGCTTCTTCATCGCATTCGGCGCAGCGATT 898  
QY 481 gcttatggccttactgtttgcgcgaacgagcagcattgatccagatgaacccgctgct 540  
|||||  
Db 899 GCTTATGGCCTTTACTTGTGTTGCCCAACGGCAGCATTTGATCCAGATGCAACCGTGTG 958  
QY 541 ccagtgctctgcaggaaacgacaaagccgaagcagaaagccgcagaaatcttcaaacgat 600  
|||||  
Db 959 CCAGTGCCTGCAGGAACGACCAAGCCGAAGCAGAGCACCAGCAGATTTCAACAGAT 1018  
QY 601 tcacacatcatccaggcacctttgacgggtgaagctattgcaactgagcggcgtcacgcat 660  
|||||  
Db 1019 TCACCATCATCCAGCACCTTTGACCGGTGAAGCTATTGCACTGAGCAGCGTCAGCGAT 1078  
QY 661 gcatgtttgcagcggaaagcttgctgcggctggtgcatgctcccaaccacgaaggcgag 720  
|||||  
Db 1079 GCCATGTTGCCACGGAAGCTTGCTCGGGCTTGCCATCGTCCCAACCAAGGGGCGAG 1138  
QY 721 ttagtcttcgggtgagtggaagattgtgtggcattcccatctggccatgcttttcgca 780  
|||||  
Db 1139 TTAGTTTCTCCGGTGAGTGAAGATTGTGTGGCATTTCCCATCTGGCCATGCTTTTCGCA 1198  
QY 781 gttcgcacaaaggctgaggatggttccaatgtggatatcttattgatgcacattggtttcgac 840  
|||||  
Db 1199 GTTCGCACCAAGGCTGAGGATGTTCCAAATGTGGATATCTTGATGCACATTTGTTTCGAC 1258  
QY 841 acagtaaacctcaacgcgcacactttaaccgctgaagaagcaggcgcatgaagtcacaa 900  
|||||  
Db 1259 ACAGTAAACCTCAACGGCAGCACCTTAAACCGCTGAAGAAGCAGGGCGATGAAGTCAA 1318  
QY 901 gcaggggagctgctgtgtaattcgatatgtatgccattaaaggctgacaggttatgaggtta 960  
|||||  
Db 1319 GCAGGGGAGCTGCTGTGTAATTGATATTTGATGCCATTTAAGGCTGCAGGTTATGAGGTA 1378  
QY 961 accacgcgatgtgttttcgaattacagaanaaacggacacctgtaaacacttacgggttgy 1020  
|||||  
Db 1379 ACCACGCCGATTTGTTTTCGAATTACAAGAAAAACGGACCTGTAAACACTTACGGTTTG 1438  
QY 1021 ggcgaattgaacgggaacacactgtccaacgtcgcaaaagaagaagcgggtgccagca 1080  
|||||  
Db 1439 GCGCAAAATTGAAGCGGAGCCAACTGCTCAACGTGCGAAAAAGAAAGCGGTGCCAGCA 1498  
QY 1081 acaccataagttgaacaccttgaggttgcg 1109  
|||||

Db 1499 ACACCATAAGTTGAACACTTGAGTGTTCG 1527  
RESULT 3  
AX127152/c  
LOCUS AX127152 349980 bp DNA PAT 11-MAY-2001  
DEFINITION Sequence 7068 from Patent EP1108790.  
ACCESSION AX127152 AX114121  
VERSION AX127152.1 GI:14041140  
KEYWORDS  
SOURCE Corynebacterium glutamicum.  
ORGANISM Corynebacterium glutamicum  
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
Actinomycetales; Corynebacterineae; Corynebacteriaceae;  
Corynebacterium.  
REFERENCE 1 (bases 1 to 349980)  
AUTHORS Nakagawa,S., Mizoguchi,H., Ando,S., Hayashi,M., Ochiai,K.,  
Yokoi,H., Tateishi,N., Senoh,A., Ikeda,M. and Ozaki,A.  
TITLE Novel polynucleotides  
JOURNAL Patent; EP 1108790-A 7068 20-JUN-2001;  
KYOMA HAKKO KOGYO CO., LTD. (JP)  
FEATURES  
Location/Qualifiers  
1..349980  
/organism="Corynebacterium glutamicum"  
/db\_xref="taxon:1718"  
/note="Seq 1 to long (3.309.400) split in 11, seq 7068  
2.700.001 3.049.980"  
BASE COUNT 81250 a 97718 c 90621 g 80391 t  
ORIGIN  
Query Match 99.9%; Score 1107.4; DB 6; Length 349980;  
Best Local Similarity 99.9%; Pred. No. 6e-297;  
Matches 1108; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 tatgatttcgggctgcagtcgagtcggtctgctcttcgggtgctgcttactaccacacgcgc 60  
|||||  
Db 110909 TATGATTTCCGGTGGTCCAGTCGGCGTCTGCTCTCGGCTGCTCTACCAATCGTC 110850  
QY 61 atcaatggtctgcacacgctccctccgcgaattgagctggagctggttttaaccagggtgga 120  
|||||  
Db 110849 ATCACTGGTCTGCACCAAGTCTCCCGCAATTGAGCTGAGCTGTTAAACAGGGTGGA 110790  
QY 121 tcttcacatctgcacacgcatctatggttaataatgcaccagggtgcgcgcatgtttggca 180  
|||||  
Db 110789 TCTTCATCTTCGCAACGGCATCTATGGCTAATATGCCCGCAGGGTGGCATGTTTGGCA 110730  
QY 181 gtgttctctgcggaagagtgtaaaagctcaaggcccttgacaggtgcttcagggtgctcc 240  
|||||  
Db 110729 GTGTTCTTCTGCGGAAGAGTGAAGAGCTCAAGGGCTTGCAGGTGCTTCAGGTGCTCC 110670  
QY 241 gctgttcttggtattacagagcctgcgatctctcggtgtgaaacctgcctgcctggccg 300  
|||||  
Db 110669 GCTGTGTTCTTGGTATTACGGAGCGCTGGCATCTTCGGTGTGAACCTTCGCTCGCGTGGCCG 110610  
QY 301 tcttcacatcggtatcggtaccgagcatatcggtggcgctttgattgacatctttaataac 360  
|||||  
Db 110609 TCTTCATCGGTATCGGTACCGCAGCTATCGGTGGCGCTTGTGATTCACCTTTTAATATC 110550  
QY 361 aaggcagttgctggtggcgctgcaggtttcttggtgtgtttctattgatgctccagat 420  
|||||  
Db 110549 AAGGCAGTTGCGTTGGCGCTGCAGGTTCTTGGGTGTTGTTCTATTATGATGCCAGAT 110490  
QY 421 atggtcatgttcttggtgtggtcagttgttaccttcttcacatgcctggcgcagcatt 480  
|||||  
Db 110489 ATGGTCATGTTCTGTTGTGTGTCAGTTGTTACCTTCTTCATCGCATTCGGCGCAGCGATT 110430  
QY 481 gcttatggccttactgtttgcgcgaacgagcagcattgatccagatgaacccgctgct 540  
|||||  
Db 110429 GCTTATGGCCTTTACTTGTGTTGCCCAACGGCAGCATTTGATCCAGATGCAACCGTGTG 110370  
QY 541 ccagtgcttcgaggaaacgacaaagccgaagcagaaagcggcgcaaatcttcaaacgat 600  
|||||



```

RESULT      5
STRSCRA     2508 bp       DNA          BCT           21-MAR-2000
LOCUS       Streptococcus mutans sucrose-6-phosphate hydrolase (scrB) gene,
DEFINITION partial cds; and enzyme scr-II gene, complete cds.
ACCESSION   M22711.1 GI:153799
VERSION     enzyme scr-II; phosphoenolpyruvate-dependent sucrose
KEYWORDS    phototransferase system; sucrose-6-phosphate hydrolase.
SOURCE      Streptococcus mutans.
ORGANISM    Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
            Streptococcus.
REFERENCE   1. (bases 1 to 2508)
AUTHORS    Sato,Y., Poy,F., Jacobson,G.R. and Kuramitsu,H.K.
TITLE      Characterization and sequence analysis of the scrA gene encoding
            enzyme IIsr of the Streptococcus mutans phosphoenolpyruvate-
            dependent sucrose phosphotransferase system
JOURNAL    J. Bacteriol. 171, 263-271 (1989)
MEDLINE    89123027
COMMENT     Draft entry and computer-readable sequence for [1] kindly provided
            by H.K.Karamitsu, 22-FEB-1989.
FEATURES   Location/Qualifiers
            source          1..2508
                        /organism="Streptococcus mutans"
                        /strain="GS-5"
                        /db_xref="taxon:1309"
                        /clone="pmH613"
                        complement(<1..28)
                        /note="sucrose-6-phosphate hydrolase (scrB)"
CDS         /codon_start=1
            /transl_table=1
            /protein_id="AAA26970.2"
            /db_xref="GI:7272364"
            /translation="MNLQNIRY"
            223..2217
            /note="enzyme scr-II"
CDS         /codon_start=1
            /transl_table=1
            /protein_id="AAA26971.1"
            /db_xref="GI:153801"
            /translation="MDYSKVASVITAVGKDNLVAHAHCATRLRLVLKDDSKVDOKAL
DKNADVKGFPTQGVQIVIGPDVAFVDEIKQTGLTEFVSDDDCLKKIAGSKKNP
IMALKLLSDIFPIIPALVAGDGMALNNFLTSEGLFGTKSLVQQFFPIIKGSSDMIO
LMSAAPFWELPIILVHSIAAKRFGANQFLASIGIMVAPGAANIIGLAANAPTSKAAT
IGAYTFWNIPLGHVTYQAASYTYQIVPVLVAVLLISLEKFHRKLPSADVFTFTPLLIS
VIITGLTIFVPVMEKYSDWLTVLWLYDTTGLMGVFGLALSPVVMTGLGHOSF
PALETOLISAFONGTGCHGDFIWTASMANVAOAGTAFAIYELTKDKMKGLSSSGVS
ALLGITPEPALFGVNLYRPFPCFALIGSASAIAQLQVVAVLSVAGSLGFLSTKA
SPIFFVCELISAFATVYGTGYKTAVDVFAEAAVEAEIEEVQEIPPEEAASAA
KAQVDEVLAAPLAFAGEVTSYNDPVFSEAMKGKAIKPSTGTVPVAPYDVQVIAF
DTGHAYGIGSDNGAEILIHIGIDTVSMEGKGFQPKVQADQKIKKGDVLGTFDSDKIAE
AGLDNTMFVTITADYASVETLASSTGTVAGVDSLLEVKK"
            2227..2252
misc_signal /note="pot. transcription termination signal; putative"
BASE COUNT 740 a 401 c 513 g 854 t
ORIGIN

Query Match              14.98; Score 165.6; DB 1; Length 2508;
Best Local Similarity    50.48; Pred. No. 5.8e-35;
Matches 495; Conservative 0; Mismatches 469; Indels 18; Gaps 3;

Qy 117 tggatcttcattcgcaacgcatctatgcgttaatatgccagagtgcgcatggtt 176
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1275 TGGTGACTTTATCTTTGTTACAGGTTCAATGGCCCATGTGCACAAGCGCAGCAACCTT 1334
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 177 ggccagtgcttcctcggcgaagagtgaaggctcaaaggccttcaggtgcgttcaggtgt 236
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1335 TGCAATTTCATTCTTAACAAAAAGATAAGAAGATGAAAGGTCCTCATCTTCTTCAGGTGT 1394
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 237 ctccgcgtttcttgttattcaggacctgcgatcttcgttgttaacccttcgcgtcgtg 296
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

Db	1395	GTACAGCGCTTCTTGCTATTACAGAACCCGCTCTCTTTTGGGGTGAAATTTGAAATATCGTTT	1454
Qy	297	gcgcttcttcacgcgtatcgcgtaccgcagcatcgcgtgctgcttgcattgcactctttaa	356
Db	1455	CCCATCTTTTGTGCTCTGATTTGGCTCAGCTAGTCTCGACGAATTCGGGGCTTACTTCA	1514
Qy	357	latacagcagtgctgcttggcgctgcaggttcttgggtgtgtttcttatgatgctcc	416
Db	1515	AGTTGTAGCGGTTCTCTCTGTTGGCTCGCAGGTTTCTTGTGTTCCCTTCTATTAAAGGCAAG	1574
Qy	417	agatatgcatgttcttgggtgtgtagtggtagcttcttcttcacgcattcgccgacgc	476
Db	1575	TTCTATTCCATTTTATGTAGTCTGTGAATTAATCAAGTTTGTCTAGCTTTTGGCAGTAAC	1634
Qy	477	gattgcttatggcctttacttggctgcgcgaagcagcatgtagcagatgcgaaccgc	536
Db	1635	TTATGGCTATGTTAGACGAAGGCTGTTGATCTCTTCTGCTGAAGCTGCTGTCTGAAGA	1694
Qy	537	tgctccagtgctgcaggaacacaaagccagacagacagaccgccgcaaatcttcaaa	596
Db	1695	AGCTATTGAAGAAGTGCAAGAAATTCACAGAGAAGCTGCTCTGTCAGCAAAATAAGGCACA	1754
Qy	597	cgattccaccatcatcc---agcgaacctttgacccgtgaagctattgcacgtgagcagcgt	653
Db	1755	AGTTACTGATGAAGTCTTTCGGCTCTCTTCTGCTGGTGAAGCTGTGTAATTAACCTCTGT	1814
Qy	654	cagcagatgccatgtttgcagcggaaagcttgctcgccgttgcacatcgctcccaaccaa	713
Db	1815	TAATGACCTGTCTTTTCTAGCGAAGCAATGGTAAGGGATTGCTATCAAACTAGTGG	1874
Qy	714	ggggcagttagttcttcgcgggtgagtggaagaattgctggtgcatctcccatctcgccatgc	773
Db	1875	TAATACAGTTTATGACCGGTTGATGGGACTGTTTCAAAATGCTTTTGATCTGGCCATGC	1934
Qy	774	tttcagcttgccaccaagggctgaggtaggttcccaatggtatcttgatgcacatgg	833
Db	1935	TTATGGTATTAACTCAGCAATG-----GTGCGGAGATTCATTATCATATTGG	1982
Qy	834	tttcacacagtaaacctcaacgcgcacgcacttcaaccgcgtgaagaagcaggcgatga	893
Db	1983	TATTGATACAGTATCAATGGAAGGTAAAGGATTGCAACAAAAAGTTCAAGCAGATCAAAA	2042
Qy	894	agtcacaaacagggggagctgctgtgtaattcgatattgatgccattaaaggctgcaggta	953
Db	2043	AATTAATAAAGGTGACGTCTCGAACAATTTGATAGTGACAAAAATTCAGAAAGCTGGTCT	2102
Qy	954	tgagtaaccacgcgattgttttgcgaattacaagaacccgacctgttaaacactta	1013
Db	2103	TGATAATACAACAATGTTCTATTGTTTACTAATACACCTGATTATGGCTCAGTTGAACCCCT	2162
Qy	1014	cggtttggcgaaattgaagcgc---ggagcacaacctgctcaacgctgcgaagaagaagc	1070
Db	2163	CGCTCTTCAGGAACCTGTTGCTGTAGGTGATAGCTTACTTTGAAGTTAAAAAATAAGAAAT	2222
Qy	1071	ggcgccagcaacaccataagtt	1092
Db	2223	ATTATCAGAAGACCGTAGGT	2244
RESULT 6			
PCR/AFOP/FC			
LOCUS	PDGRAFOPER	21838 bp	DNA
DEFINITION	Pediococcus pentosaceus raffinose operon genes.	BCT	23-MAY-1994
ACCESSION	L32093		
VERSION	L32093.1	GI:493181	
KEYWORDS	agar gene; agaS gene; agl gene; alpha-galactosidase; alpha-glucosidase; fructokinase; insertion element; permease; rafp gene; rafR gene; raffinose operon; regulatory protein; scrA gene; scrB gene; scrK gene; scrR gene; sucrose-6-phosphate; transport protein.		
SOURCE	Pediococcus pentosaceus (strain PPE1.0) DNA; Insertion sequence IS30 homolog (transposable element insertion sequence IS30 homolog, kingdom Prokaryotes) DNA; Pediococcus pentosaceus (strain PPE1.0)		

DNA; Insertion sequence IS3 homolog (transposable element Insertion sequence IS3 homolog, kingdom Prokaryotae) DNA; and *Pediococcus pentosaceus* (strain PPE1.0) DNA.

**ORGANISM** *Pediococcus pentosaceus*  
Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae; *Pediococcus*.

**REFERENCE** 1 (bases 1 to 21838)  
**AUTHORS** Leenhouts, K.J., Bolhuis, A.A., Kok, J.J. and Venema, G.G.  
**TITLE** The sucrose and raffinose operons of *Pediococcus pentosaceus* PPE1.0  
**JOURNAL** Unpublished  
**COMMENT** On May 25, 1994 this sequence version replaced gi:475106.  
**FEATURES**  
    **source**  
        1. 21838  
            Location/Qualifiers  
            /organism="Pediococcus pentosaceus"  
            /strain="PPE1.0"  
            /db\_xref="taxon:1255"  
            487..510  
            /notes="DR2; putative"  
            /rpt\_type=direct  
            488..561  
            /notes="IR1; putative"  
            /rpt\_type=inverted  
            513..561  
            /notes="DR1; putative"  
            /rpt\_type=direct  
            2724..2760  
            /standard\_name="terminator 4"  
            /notes="putative"  
            complement(2929..3864)  
            /genes="rafr"  
            complement(2929..3762)  
            /genes="rafr"  
            /function="regulation raffinose operon"  
            /notes="putative"  
            /codon\_start=1  
            /transl\_table=11  
            /product="regulatory protein"  
            /protein\_id="AA25562.1"  
            /db\_xref="GI:475107"

repeat\_unit

repeat\_unit

repeat\_unit

terminator

gene

CDS

/translation="MGEYKTLANKSFESNVLFEGQACLPNYYKGNVRSYVYIH  
IQGKFTFAANHPATVLRAGDIFILPKGTPCYQADNDQPKFYWGFSAGIRIEM  
LSGLLAQYKQVQNGHIYADLSLYVLHPNLSINDVLGSLIYLFYDLSRE  
PADATNIKYKTEQFNLAWSYLOENYSTGCTIMDLCHYLNLSRYLYTLFKHTANTSP  
QKLLTKLRLEDAKQRLSTNNNSVQSIANVMGVYKDSFTFSKAFKRYSGASPSYRKSG  
I"  
complement(3837..3842)  
/genes="rafr"  
complement(3859..3864)  
/genes="rafr"  
3909..5918  
/genes="rafr"  
3909..3914  
/genes="rafr"  
/notes="putative"  
3932..3937  
/genes="rafr"  
/notes="putative"  
3993..5918  
/genes="rafr"  
/function="raffinose transport protein"  
/notes="putative"  
/codon\_start=1  
/transl\_table=11  
/product="permease"  
/protein\_id="AA25563.1"  
/db\_xref="GI:475108"

-10\_signal

-35\_signal

gene

-35\_signal

-10\_signal

CDS

KFEGALSNGVVGQIAISGMWTTGATASSITAGQALHFKLTMFAPFALLMIIATGIFSK  
QIFLTEKIAEIVAELETRWTKFDNTDQVAKVVTSLDLATPIAGQVPLAQVNDP  
TAAAGTGDGEFAIKPSDGRILAPFDATVQVETTRAVGLVGDNGIIVLHIGLGVK  
LKGTFISYVEEQHQVQGDLELLEFWDPTIKAGLDDTVIMVTNSTETFTMDWLKRP  
GQAVKATDNLQLHTRKA"  
5976..8177  
/genes="agar"  
5976..8177  
/genes="agar"  
/codon\_start=1  
/transl\_table=11  
/evidence=experimental  
/product="alpha-galactosidase"  
/protein\_id="AA25564.1"  
/db\_xref="GI:475109"

gene

CDS

/translation="MSLIIVDQANRVFHLHNQTLSYIFAVEQGGTSLHLYFGGHVDHY  
HGLRYPRVDRGFSGNLPSTDRFTSLPKREYSTAGENDYHLPAAIVRHTDGANAL  
YLVYQGYRIEAGKPKLSGLPAFVDETEAETLIVLVQVQSVFEDLQYTIYDRPV  
VTRSVQVCNGQHAVNLEKVASQIDFTDQFETITLPGAHANEHPERGSINYGQT  
FGSLRGTSSHQNPFLALVDHTTTFESGDAYGENLYSGNHAPELKDOLDLHLAVG  
INSYNEFWOLKAGATQTPVELMVYTNKGLNMSOAYHHLIRERVSRSEKQDERPIV  
VANWEATPFDFNEAKLPIVDEAKQLGIEMFVLDDGFWGHRDDDNSSLGQWQVDHRKF  
PQGLNFHVYVHEQGLKFGIWEPEMISYDSKLIYQQHPDILMQVGRSPSRNQIIL  
DLGRQAVRNIFDQLLKSQIDYIKWDMNRHLSDIYSVALPPRQGEVHYRVLG  
LYELLERLTAYPHILFEGCGGGGFRDAGMAYMPQIWASONTDAVALTIQYGTSL  
AYPISLATAHVSVPNQGTGRTSMSTRSAVASGLVGLYELDLTOLSSADKOIVQKV  
VOYKOIRPLIQEGEFYRLKSPITSNOAMFVSPQODEALVMVFNLTSAQPSLTNKT  
LVGLNPKLNYQNIATKAIFGGDELMQLGFDVPVYQDYTKYVHFKAVENT"  
8265..8289  
/standard\_name="terminator 3"  
/notes="putative"  
9403..9426  
/notes="DR2; putative"  
/rpt\_type=direct  
10396..10442  
/notes="DR1; putative"  
/rpt\_type=direct  
10397..10443  
/notes="IR2; putative"  
/rpt\_type=inverted  
10881..10716  
/standard\_name="terminator 2"  
/notes="putative"  
complement(10727..11593)  
/genes="scrK"  
complement(10727..11593)  
/genes="scrK"  
/function="fructose 6-phosphotransferase"  
/notes="putative"  
/codon\_start=1  
/transl\_table=11  
/product="fructokinase"  
/protein\_id="AA25565.1"  
/db\_xref="GI:475110"

terminator

repeat\_unit

repeat\_unit

repeat\_unit

terminator

gene

CDS

gene

CDS

/translation="MLLGAIEAGGTKEVCATGAENGQVSDRISIPPTTTPVETMTAVDD  
YFTHPVDAIGISFGPIGVNPHDKYGIITTPKPGWGDQFLGHLKQFNLPLYVT  
TDVNEAAYGESMIGIAKDPNSIYMTIGTGVGAGVISQNHIFNGRHTLGHMRLNRL  
PGDDFKSNCPYHDICLEGLAGPAGKRTGKAGKIDPDDPWPPIITDYIAQACVNL  
VAFAPDKIILNGVMNQRLFPFIREKFAAYLNGYEEVPLDDIIVPAGLGNNSGIAT  
GLLQAALKNA"  
complement(11673..13832)  
/genes="agas"  
complement(11673..13832)  
/genes="agas"  
/notes="putative"  
/codon\_start=1  
/transl\_table=11  
/product="alpha-galactosidase"  
/protein\_id="AA25566.1"  
/db\_xref="GI:475111"  
/translation="MPVEYDPKTLNLHNDQISYVIQILAHRYPVHRVYFGRVFSKQP  
YFEPMPGSHAFANDPTEPFYVSLPSEYSTIGSGDYRQPAYVTKDANNQLLPFILE  
YTCFSYNDQPIINSRQLPPTVPTTTLVIHLTDAVTKLQMDLNTTIFENQPLILRS

TTLRHGGTTNLQVLTALSSAQIDLPTDOYTAITLSTGTHAEANFSNRLHPGLQTVKSL  
RTGSPQSPFMALAEPTTELAGTVIGCALWAGSNFDTEVDVQYQHSRLTIGLEPTL  
TFEQLQKSPNSFOTPEALVTITNGMSQVDFSYQLMPQSOTIPNWTMETLT  
FAYSEQVQHLIEHAHQGLQMLVDDGNGVNRNGENGLQDFWFDVPKFPNGNLPLA  
QQAHHHRMKFGLTVPEPMITNSQLQWQHPDWLVQVVDRTPIATRIQLVLDLSQAQAVR  
DLH1TUTLNVQNNOLDYKWDMMNRLHTQVGSITLPAAOQGLYHRVVCGLYLDLRL  
KRACPGLIENCSAGGRDFGMLPTYNQWISDLTDPVDVDRATIEGFSYLPDPRIFS  
NH1TASPNQAGRTTPEPTELQACLGQLELNPQLAPSEQQLRGALLKKYQQLKS  
TF1KAHYFRTTHRVVWLVITADNKQAICCCNLNGNSRVKTKQHPHLPULHYDLAEAY  
SDSSGNRYTHQGNMTMGIPUKPNADFTSOLIYLCON"

gene  
complement(13852..15989)  
/gene="scra"

CDS  
complement(13852..15807)  
/gene="scra"  
/function="enzyme II of the PTS system, sucrose specific"  
/note="enzyme IIabc"  
/codon\_start=1  
/transl\_table=11  
/evidence=experimental  
/protein\_id="AAA25567.1"  
/db\_xref="GI:475112"  
/translation="MNHQEVADRVLNAIGNKNITQAAAHCACTRLRLVKDESKDQCAL  
DDADVKGTPTNGQVYIIGPGVDKVDYDALIVKTKGLEKVTDDIKVANAAGCONKP  
LMDFLKVLSDFIPIPALVAGGLLNALNNVLTAEHFLMAKSVVEYVPLGKGAEMIN  
AMASAPTEPILLEGFSATRRFGGCPYLGATGMIMVPLSVNGSVATTMAAGKRVY  
WNVFLGHAQAQGOQVLPVGLVAFILATLEKFMHFKGAFDTPTMFAIVITGFL  
TFTITGVPRLTSDGLNTGLVNTGTGMGIFGLYSIVITGLHDTFPAIETQL  
LANVAKTGGSFIPVASMANGIGOCATLAIFATKSKOKALTSSAGVSALLGITPQL  
IFGVNLKMKTPPEVFAIASGASAFGLFHLVLSVAMPASVIGFISTASKSIPAFMLS  
AVTSVFVAIPTFLYAKRTIGDDRDVKSPATSTVNVNDEILISAPVGAESLSQV  
NDQVFSAEIMCKGNAIYVPSDSQVVPADGVITVYDSGHAYIGDKTAPGABILLHGLD  
KVTNVQAGEQLVALTAPAASSVAATT"  
complement(15872..15877)  
/gene="scra"  
complement(15894..15989)  
/gene="scra"  
15993..15998  
/gene="scrB"  
15993..17568  
/gene="scrB"  
16017..16022  
/gene="scrB"

gene  
complement(15872..15877)  
/gene="scra"

-10\_signal  
complement(15894..15989)  
/gene="scra"

-35\_signal  
15993..15998  
/gene="scrB"

-35\_signal  
15993..17568  
/gene="scrB"

gene  
complement(15872..15877)  
/gene="scra"

-10\_signal  
complement(15894..15989)  
/gene="scra"

Query Match	12.5%	Score 138.4	DB 1	Length 21838
Best Local Similarity	57.9%	Pred. No. 2.5e-27		
Matches	272	Conservative 0	Mismatches 186	Indels 12
Gaps	1			
32	tcttcggctgtgctactaccacatcgtctatcactcgtctgctgcacagctccttcgcgcga	91		
14879	TCTTTGGTTTATATATCTGCAAATGTTATTACTGTCTCCATCAAACTTTTCCACGAA	14820		
92	ttgagctggagctgtttaaacag-----ggtggactcttcactcttcgcgaacg	139		
14819	TCGAAACCCAGTCTGTGGCCAAATGTTGCCAAAACCTGGTGGTCTCTTTATCTTCCCA	14760		
140	catctatggctaatatcgccacaggtgcggcatgtttggcagtgctcttcctgcggca	199		
14759	CCTCGATGGCCAAACATTTGGGCAAGGTGCCGCCACTTTAGCTATTCTTTGCCACTA	14700		
200	gtgaaaagctcaaggccttgcagggtgtcagggtgtctcgcgtgtctcttggtattac	259		
14699	GCCAAAAGCAAAAAGCCCTGACTTCTTCAGCCGGGGTCTCAGCGGTACTCGGAAT	14640		
260	agcctcgcatcttcgggtgtaaaccttcgcctgcctgcctgccttcctcatcggtatc	319		
14639	AACCTGCTATTTTCGGGGTCAACCTCAAGATGAAATTTCCATTTGTCTTTGACGCG	14580		
320	ccgcagctatcgggtggcgcttctgattgcactctttaaatacaaggcagctgcgctgg	379		
14579	CCTCAGGAATTGCTTCAGCTTCTTAGGACTATTCACAGTTCTGTCCTCGTGGCGAT	14520		
380	ctgcaggtttcttgggtgtgttcttattgatctccagatattggtcatgttcttggtgt	439		

Dbb\_14519 CCGCATCAGTCATGGGATTCAATTTCGATGTCTTCAAAGTGCAATGCCACGCATTCATGCTCA 14160

Qy      440 gtcgagttgttaccttctccatcgcattcgccggcaggagtcttgcttatgcc 489  
||||| | | | | ||||| | | | | |||||  
Dbb\_14459 GTGCTGTCATCTCGTTCCGTCGTCGAATTAFTCCAACCTTTTATCATGCCC 14410

RESULT 7  
PPSURFOP/c LOCUS PPSURFOP 21839 bp DNA BCT 24-MAY-1994  
DEFINITION P.pentosaceus (PPEI.0) sucrose and raffinose operons.  
ACCESSION Z32771  
VERSION Z32771.1 GI:493728  
KEYWORDS alpha-galactosidase; alpha-glucosidase; enzyme IIabc; fructokinase;  
insertion element; permease; regulator; sucrase.  
SOURCE Pediococcus pentosaceus  
ORGANISM Pediococcus pentosaceus  
Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;  
Pedococcus.

REFERENCE 1 (bases 1 to 21839)  
AUTHORS Leenhouts,K.K.J., Bolhuis,A.A., Kok,J.J. and Venema,G.G.  
TITLE The sucrose and raffinose operons of *Pediodoccus pentosaceus* PPEI.0.  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 21839)  
AUTHORS Leenhouts,K.K.  
TITLE Direct Submission  
SUBMITTED (27-APR-1994) Leenhouts K. K., Institute for Biological Sciences,  
Department of Genetics, Kercklaan 30, Haren, The Netherlands, 9751 NN  
COMMENT On May 26, 1994 this sequence version replaced gi:475962.  
FEATURES location/Qualifiers  
source .1..21839  
          /organism="Pediococcus pentosaceus"  
          /strain="ppeI.0"  
          /db\_xref="taxon:1255"  
repeat\_unit 487..510  
             /note="DR2"  
             /citation=[1]  
             /rpt\_type=DIRECT  
repeat\_unit 488..561  
             /note="IRI"  
             /citation=[1]  
             /rpt\_type=INVERTED  
repeat\_unit 513..561  
             /note="DRI"  
             /citation=[1]  
terminator 2724..2760  
             /standard\_name="terminator 4"  
             /citation=[1]  
gene complement(2929..3864)  
             /gene="rafr"  
CDS complement(2929..3762)  
             /gene="rafr"  
             /standard\_name="regulator"  
             /function="regulation raffinose-operon"  
             /citation=[1]  
codon\_start=1  
transl\_table=11  
product="regulator"  
protein\_id="CAA83663.1"  
db\_xref="GI:475963"  
db\_xref="SWISS-PROT:P43465"

-10\_signal translation="MNGEYKTLANKSFESNLFVEGDACLPNITYKGNNVRSVVHY IQSGKTFAAHPAVLKAGDFILPKGTPCYQNDDPWKYFWLGFSGARIEAM LGSLAQKCYLROVGNGHIYADLSLYKVHLPNSLDIVLGISLIYRLFYDLRLM PADATNIKVSQTQCNLAWSYLQENYSTGCTIMDLCHYNLSRYLTFLFKTHANTSP QKLTKLRLEDAQRLLSTSNNSQSIANMVGYKDSTTFSKAFKRYSGASPSYYRKSIG I"

-35\_signal complement(3837..3842)  
             /gene="rafr"  
complement(3859..3864)  
             /citation=[1]



Db 14819 TCGAAACCAAGTGTGTGGCCCAATGTTGCCAAACCTGGTGGTTCCTTTATCTTCCAGTCG 14760  
Qy 140 catctatcgctaataatcccccagggtgcggcatgtttggcagtgcttctctcctggcgaaga 199  
Db 14759 CCTCGATGCCCAACATTTGGGCAAGGTGGCGGCACATTTAGCTATTCTTTGTCACCAATAAGA 14700  
Qy 200 gtgaaaagctcaaggcccttcaggtgcttcaggtgtctccagtgcttcttggtattacgg 259  
Db 14699 GCCAAAACCAAAAGCCCTGACTTCTTCAGCCGGGGTCTCAGCGTTACTTCGGAATACGG 14640  
Qy 260 agcctgcgactctcggtgtgaaccttcgctgcgctgcggcttcttcacgtatcggtacgta 319  
Db 14639 AACCTGCTATTTCGGGTCAACCTCAGATGAATTTCCATTTCCTTCAGCGGATG 14580  
Qy 320 ccgcagctcgttggtgcgcttgattgacactctttaataatcaaggcagttcggtggcg 379  
Db 14579 CCTCAGGAATTCGCTTCAGCTTTCTTAGGACTATTCCAGGTTCTGTCGCGTGGCGATGGGTC 14520  
Qy 380 ctgcaggtttctgggtgtttcttctattgatgtccagatattggtcatgttcttgggtg 439  
Db 14519 CCGCATGCTATCGATCGATTCGATTTCGATTGCTTCAAAGTCGATCCAGCATTCATGCTCA 14460  
Qy 440 gtgcaggtgttacccttcttcctcgcattcgcgcgagcgatgcttatggc 489  
Db 14459 GTGCTGCTATCGTTCGTCGCGATTTATTCACATTTTATCTATGCC 14410

RESULT 8  
AP001508 296950 bp DNA BCT 10-JAN-2001  
LOCUS Bacillus halodurans genomic DNA, section 2/14.  
DEFINITION AP001508 BA000004  
ACCESSION AP001508.1 GI:10172890  
VERSION AP001508.1  
KEYWORDS  
SOURCE  
ORGANISM  
Bacillus halodurans DNA.  
Bacillus halodurans  
Bacteria; Firmicutes; Bacillus/Clostridium group;  
Bacillus/Staphylococcus group; Bacillus.  
1 (sites)  
Takami,H. and Horikoshi,K.  
Reidentification of facultatively alkaliphilic Bacillus sp. C-125  
to Bacillus halodurans  
Biosci. Biotechnol. Biochem. 63, 943-945 (1999)  
2 (sites)  
Takami,H., Takaki,Y., Nakasone,K., Sakiyama,T., Maeno,G.,  
Sasaki,R., Hirama,C., Fuji,F. and Masui,N.  
Genetic analysis of the chromosome of alkaliphilic Bacillus  
halodurans C-125  
Extremophiles 3 (3), 227-233 (1999)  
9411980  
3 (sites)  
Takami,H., Nakasone,K., Hirama,C., Takaki,Y., Masui,N., Fuji,F.,  
Nakamura,Y. and Inoue,A.  
An improved physical and genetic map of the genome of alkaliphilic  
Bacillus sp. C-125  
Extremophiles 3 (1), 21-28 (1999)  
99184645  
4 (sites)  
Takami,H., Masui,N., Nakasone,K. and Horikoshi,K.  
Replication origin region of the chromosome of alkaliphilic  
Bacillus halodurans C-125  
Biosci. Biotechnol. Biochem. 63 (6), 1134-1137 (1999)  
99356711  
5 (sites)  
Takami,H., Takaki,Y., Nakasone,K., Hirama,C., Inoue,A. and  
Horikoshi,K.  
Sequence analysis of a 32-kb region including the major ribosomal  
protein gene clusters from alkaliphilic Bacillus sp. strain C-125  
Biosci. Biotechnol. Biochem. 63 (2), 452-455 (1999)  
99209008  
6 (sites)  
Takami,H.

TITLE Genome analysis of facultatively alkaliphilic Bacillus halodurans C-125  
JOURNAL (in) Extremophiles in deep-sea environments (Ed.);  
HORIYOSHI, K. TSUJII;  
: 249-284; Springer-Verlag (1999)  
REFERENCE 7 (sites)  
AUTHORS Takami,H., Nakasone,K., Ogasawara,N., Hirama,C., Nakamura,Y.,  
Masui,N., Fuji,F., Takaki,Y., Inoue,A. and Horikoshi,K.  
Sequencing of three lambda clones from the genome of alkaliphilic  
Bacillus sp. strain C-125  
Extremophiles 3 (1), 29-34 (1999)  
99184646  
8 (sites)  
Takami,H. and Horikoshi,K.  
Analysis of the genome of an alkaliphilic Bacillus strain from an  
Industrial point of view  
Extremophiles 4 (2), 99-108 (2000)  
20263314  
REFERENCE 9 (sites)  
AUTHORS Nakasone,K., Masui,N., Takaki,Y., Sasaki,R., Maeno,G., Sakiyama,T.,  
Hirama,C., Fuji,F. and Takami,H.  
Characterization and comparative study of the rrr operons of  
alkaliphilic Bacillus halodurans C-125  
Extremophiles 4 (4), 209-214 (2000)  
20426005  
10 (sites)  
Takami,H., Nakasone,K., Takaki,Y., Maeno,G., Sasaki,R., Masui,N.,  
Fuji,F., Hirama,C., Nakamura,Y., Ogasawara,N., Kuhara,S. and  
Horikoshi,K.  
Complete genome sequence of the alkaliphilic bacterium Bacillus  
halodurans and genomic sequence comparison with Bacillus subtilis  
Nucleic Acids Res. 28 (21), 4317-4331 (2000)  
20512382  
REFERENCE 11 (bases 1 to 296950)  
AUTHORS Takami,H. and Takaki,Y.  
TITLE Direct Submission  
JOURNAL Submitted (22-MAR-2000) Hideto Takami, Japan Marine Science and  
Technology Center, Deep-sea Microorganisms Research Group; 2-15  
Natsushima, Yokosuka, Kanagawa 237-0061, Japan  
(E-mail: takami@jamstec.go.jp/jamstec-e/bio/DEEPSTAR/Research.html,  
URL: http://www.jamstec.go.jp/jamstec-e/bio/DEEPSTAR/Research.html,  
Tel: 81-468-67-3895, Fax: 81-468-66-6364)

FEATURES  
source  
1..296950  
/organism="Bacillus halodurans"  
/db\_xref="taxon:86665"  
/note="alkaliphile"  
complement(427..615)  
/gene="BH0278"  
complement(427..615)  
/gene="BH0278"  
/note="BH0278"  
unknown  
/codon\_start=1  
/transl\_table=11  
/protein\_id="BAB03997.1"  
/db\_xref="GI:10172891"  
/translation="MRKMGTLVLASIFLVACNADENETESVDQNAEEVEEVEDELE  
QDQEEFDVVEETIDDEE"  
798..1163  
/gene="BH0279"  
798..1163  
/gene="BH0279"  
/note="BH0279"  
unknown  
/codon\_start=1  
/transl\_table=11  
/protein\_id="BAB03998.1"  
/db\_xref="GI:10172892"  
/translation="MKIFFVAAATFLAASQQAALAEPTVSEQGVSNADVTTFE  
TVDAQSTWVWSRGITNTDTKRTAYSNYHRSAPHSAAQLGNDRAVSPGLGPGATT  
YASIRGNKLNLPATFVFY"  
1334..3475  
gene





Db	20677	TTGATGTCAGCGCAATTACAGCAGCGGGTATGATGATCATCAGCGCTGTTCTGATTACAA	20736
QY	983	attacaagaaacccgacctgtaaacacttacggtttggcgaaattgaagcgggagccca	1042
Db	20737	ATGCAAAACAATTTTCAATGTACAAACGACAGATAAGCGCAGGTGCACAAAGTGAGGATT	20796
QY	1043	acctgctcaactcgcaaaagaa	1065
Db	20797	TGCTCATTCACGTCATAAAGTAA	20819
RESULT	9		
LOCUS	SXSXCA	2655 bp	DNA
DEFINITION	S.xyloosus scrA gene and unidentified open reading frames.		
ACCESSION	X69800		
VERSION	X69800.1	GI:407905	
KEYWORDS	membrane protein; scrA gene; sucrose transport protein.		
SOURCE	Staphylococcus xyloosus.		
ORGANISM	Staphylococcus xyloosus.		
REFERENCE	Bacteria; Firmicutes; Bacillus/Clostridium group;		
AUTHORS	Bacillus/staphylococcus group; Staphylococcus.		
TITLE	Wagner, E.		
JOURNAL	Submitted (15-DEC-1992) E. Wagner, Mikrobielle Genetik, Universitaet Tuebingen, Auf der Morgenstelle 28, 7400 Tuebingen 1, FRG		
REFERENCE	2 (bases 1 to 2655)		
AUTHORS	Wagner, E., Gotz, F. and Bruckner, R.		
TITLE	Cloning and characterization of the scrA gene encoding the sucrose-specific Enzyme II of the phosphotransferase system from Staphylococcus xyloosus		
JOURNAL	Mol. Gen. Genet. 241 (1-2), 33-41 (1993)		
MEDLINE	94049686		
FEATURES	Location/Qualifiers		
source	1..2655		
	/organism="Staphylococcus xyloosus"		
	/strain="DSM 20267"		
	/isolate="cured strain C2a"		
	/db_xref="taxon:1288"		
	/clone_lib="sub-genomic in Staphylococcus xyloosus mutant scr1"		
	/clone="pSE4"		
	complement(1..121)		
CDS	/note="ORF3"		
	/codon_start=1		
	/transl_table=11		
	/protein_id="CAA49459.1"		
	/db_xref="GI:407906"		
	/db_xref="SPTREMBL:Q56203"		
	/translation="MKSTLTCIHEFLNTHQRCIHQVILLFLSLNKLDTVLTNGY"		
RBS	complement(132..137)		
RBS	338..342		
CDS	350..925		
	/note="ORF2"		
	/codon_start=1		
	/transl_table=11		
	/protein_id="CAA49460.1"		
	/db_xref="GI:407907"		
	/db_xref="SPTREMBL:Q56204"		
	/translation="MQHIYSEFYNOESKITKLKSVGLIVLHDDVNNNSAEDIYVWLQQRIDNNELEKAWACYVDKQCYWPHSPYVEWHGCGNAPFANDHYIGIERCQSKINGILTDQGEENEASWTAAALLKKYQLPVNRVETVKLHKMFDDTECPARAWSIHLDNAPTNBENIKMLQDYFDIKIKYHFVNKESDMFVVG"		
	-35_signal		
	-10_signal		
RBS	995..1000		
gene	1040..1045		
	1053..2495		
	/gene="scrA"		
	1053..2495		
CDS	/gene="scrA"		

/note="ORF2"			
/codon_start=1			
/transl_table=11			
/product="EiIsr"			
/protein_id="CAA49461.1"			
/db_xref="GI:407908"			
/db_xref="SWISS-PROT:P51184"			
/translation="MNYKSAENILQALGGEDNVAMTHCATRLRLVLKDEGLVDEKALGDMDVVKFTSTGGYQVILGGSTGVNKKVSELEKITGKEASVSVKTKGNMNPFRQFVKMSDFIVPIIPAIVAGGLMGINNLTAPGIFYDNQSLIEVQNGFSGLAKEIIPANAPFTLLPILIGESAARFGGNAYLGAALMILVHPMLMSAYDPTPLAALSTGFHWNLFGLNEQVGOVLPMVATVILATIEKLRKVPTVLDNLLTPLLAALSTGFITFSFVGLTRTLGVLSDGLTWLVEFGAIGGLIFGLLYAPITVTGMHHSFIATETQLIADSSSTGGSFIFPIATMNSIAQGAALAAAFIIKENKKLKGVAASGVSALLGITEPMFGVNLKLRYPFIAGVSGISGAYIAFFKFKKALATAGTIPGISISGQNGWMLHYGIAMTIAFIVAFGVYALSRYKKRNIEA"			
BASE COUNT	909 a	347 c	521 g
ORIGIN	878 t		
Query Match	12.2%	Score 135.2;	DB 1; Length 2655;
Best Local Similarity	60.1%	Pred. No. 1.7e-26;	
Matches	268;	Conservative 0;	Mismatches 163; Indels 15; Gaps 2;
QY	1	tatgatttcggcggtccagtcgcgcgtctcttcggtctggtctactcaccacgcgc	60
Db	1953	TATGAATTTGGTGGTGTATCGGTGGCTTATATTTGGTTTATATGCGCCCAATCGTT	2012
QY	61	atactgggtctgcaccagtccttcctccgcgaattgagctggagctgtttaacca-----	113
Db	2013	ATTACAGGTATGCATCATAGTTTTATTGCTATTGAAAGCGCACTTATCGCTGACAGTTCA	2072
QY	114	-----gggtggatccctcatcttcgcgaacggcatctctatggctaatatcgccaggggtgcg	168
Db	2073	TCAACGGGTGGTTTCATTATATATCCCTATTGCAACAATGCTAATATTGCACAGGAGCG	2132
QY	169	gcagtgttgccagctgttcctcctggcga---agagtaaaagctcaagggccttcgaggt	225
Db	2133	GCTGCGCTGCACGGTCTCTTTATTATAAAGAAATAAATAAAGGCGTAGCCTCT	2192
QY	226	gcttcagggtctcccgctgcttcttggtattacggagcgtcgatcttcggtgtaacctt	285
Db	2193	GCTGAGGTGTATCAGCTCTATTAGGTATTACAGAACCACTGTTTGTGTGTAAGTTG	2252
QY	286	cgcctgcgtgcgttccttcctcatcggtatcggtacgcagctatcggtgcgcttgatt	345
Db	2253	AAGCTAAGATATCCATTATATAGGGCAATCGTAGGTTTCAGGTATTGGTCTGCTATATA	2312
QY	346	gcactctttaataatcaagcagttgcttggtggcgctgcaggtttcttggtgtgtttct	405
Db	2313	GCATTCTTTAAGGTTAAAGCTATTGCATTAGGTAGTACTGCAGGTATACCAAGGATTATATCT	2372
QY	406	attgattgctccagataggtcatggtt	431
Db	2373	ATTAGTGGTCAAAATAATGTTGGTT	2398
RESULT	10		
LOCUS	LLZ97015	5800 bp	DNA
DEFINITION	Lactococcus lactis cremoris sucrose gene cluster.		
ACCESSION	Z97015		
VERSION	Z97015.1	GI:4581474	
KEYWORDS	enzyme II sucrose protein; fructokinase; sacA gene; sacB gene; sack gene; sacR gene; sucrose-6-phosphate hydrolase.		
SOURCE	Lactococcus lactis.		
ORGANISM	Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae; Lactococcus.		
REFERENCE	1 (bases 1 to 5800)		
AUTHORS	Luesink, E.J., Marugg, J.D., Kuipers, O.P. and de Vos, W.M.		
TITLE	Characterization of the divergent genes sacB and sacAR operons, involved in sucrose utilization by Lactococcus lactis		
	05-APR-1999		



Qy	415	ccagatatggtcatgttcttgggtgtgtgagtcagttgattacattcttctcatcgcatcgccga	474
Db	4212	GGGTATAACTTACAAATTTATGATTTTCGATATTTATTTAGCTTCTTCTTATTTGCTTCTTCTT	4271
Qy	475	gcgattgttgccttcttactgttgcgcgcaacgagcagcattgattccagatccaacc	534
Db	4272	ACCTCAATCTATGTCGCGGATGGAACCTAAATCTATTACGAAGAAATAAATACAG	4331
Qy	535	gctgctccagtcgctgaggaacacacaaagcagagaagcagccgagcaatttca	594
Db	4332	AATGC-----AACAACTCAATACCAACCTGAGAAAGTTATTATCGATCCA	4376
Qy	595	aacattccaccatcatccagccaccccttgaccggtgaagctattgcaactgagcagcgc	654
Db	4377	GTTAAAGTGGCAACTCTTCTGCTCCGATAAATGATTGTTGATTCCTCTGCTGATGTA	4436
Qy	655	agcattgcatgttgcagcggaagcttgcctgctggcgcttgccattgctcccaaccaag	714
Db	4437	AGTGACCCCTGTTTCTCAAAAGAAATATTGGGAAAGCTATTGCAATCAAGCCTAAATCT	4496
Qy	715	ggcaggttagtttcccggtgagtggaagattgtggtggcattcccatctggccatgct	774
Db	4497	GGAAACATTTTCTCACCAGCGACGGAGAAATATTATTGCTTATGAAACAGGTCACGCT	4556
Qy	775	ttccagttcgcacaaagctgagatggtttccaatgtggatatcttgatcacattgct	834
Db	4557	TATGTTATAAAACAAAATGGAGGT-----GAAGTCTCTTTACATATTGGA	4604
Qy	835	ttcacacagtaaacctcaacgagcagcacttcaaccgcctgaagaagcagggcgatgaa	894
Db	4605	ATAGATACTGTTTCAATGAATGTTAATGATTATCATAAAATGTTAAAGTTGGCCAGAA	4664
Qy	895	gtcaaacagggagctgctgtgaaatctgatatgtgattgcatgaagcgtgcaggttat	954
Db	4665	GTAAGACGAGGGATTTTACTAGGATCTTTTGTATAAGAGAAGAAATCAAGAAGATGCGATTG	4724
Qy	955	gaggtaacccgagctgtgtgttctcgatt	985
Db	4725	GATGATCTGTAATTTGTTATTACAAT	4755
RESULT	11		
AE004395/c			
LOCUS	AE004395	10264 bp	DNA
DEFINITION	Vibrio cholerae chromosome II, section 52 of 93 of the complete chromosome.	BCT	31-JUL-2000
ACCESSION	AE004395		
VERSION	AE004395.1		
KEYWORDS	GI:9658068		
SOURCE	Vibrio cholerae.		
ORGANISM	Vibrio cholerae		
REFERENCE	Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.		
AUTHORS	Heidelberg, J.F., Eisen, J.A., Nelson, W.C., Clayton, R.A., Gwinn, M.L., Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Umayam, L., Gill, S.R., Nelson, K.E., Read, T.D., Tettelin, H., Richardson, D., Ermolaeva, M.D., Vamathevan, J., Bass, S., Qin, H., Dragoi, I., Sellers, P., McDonald, L., Utterback, T., Fleishmann, R.D., Nierman, W.C. and White, O.		
TITLE	DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae		
JOURNAL	Nature 406 (6795), 477-483 (2000)		
MEDLINE	20406833		
REFERENCE	2 (bases 1 to 10264)		
AUTHORS	Heidelberg, J.F., Eisen, J.A., Nelson, W.C., Clayton, R.A., Gwinn, M.L., Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Umayam, L.A., Gill, S.R., Nelson, K.E., Read, T.D., Tettelin, H., Richardson, D., Ermolaeva, M.D., Vamathevan, J., Bass, S., Qin, H., Dragoi, I., Sellers, P., McDonald, L., Utterback, T., Fleishmann, R.D., Nierman, W.C., White, O., Salzberg, S.L., Smith, H.O., Colwell, R.R., Mekalanos, J.J., Venter, J.C. and Fraser, C.M.		
TITLE	Direct Submission		

JOURNAL	Submitted (14-JUN-2000) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
FEATURES	Location/Qualifiers
source	1. .10264
	/organism="Vibrio cholerae"
	/strain="Ni6961"
	/serotype="O1"
	/db_xref="taxon:666"
	/chromosome="II"
gene	/note="biotype: El Tor"
	complement(114. .1553)
CDS	/gene="VCA0653"
	complement(114. .1553)
	/gene="VCA0653"
	/note="similar to GB:M76768 SP:P22825 PID:155263; identified by sequence similarity; putative"
	/codon_start=1
	/transl_table=11
	/product="PTS system, sucrose-specific IIBC component"
	/protein_id="AAF96554.1"
	/db_xref="GI:9658069"
	/translation="MDIPVIAKQLLESGLGKSNQALAHCASTRRLRLVNDLQINESAIESQGVKQKPVAGQYQIIFSGIVNQVYAEAMKLTGLVEMSTNDVASAAGKQNWQRAVGLSDIFVRIIPAIVAGLLMFLNLTAFGLTEGSLDANGLADLASMINTFANAPFVLPVLLAFASRKFQGNPFLGAALGMVHPDLINGWFGSASVSGTPTWNIGFEIEKVGYSVLPLVSAILAKIENGLRKIVPSVVDNLLTTHLAFITGFLTFTVVGPLTRDVGFWLGDALNWLVDYSGFVGALGFIYAPFVITGMHHSIATLQTLADITGCTGTFIIPAAAMNSIAQGAALAVGMKTKLKGVAIPSGVTALLGITEPAMAIISFAAFALTLLAMREKTQAQVTA"
gene	1854. .2846
CDS	/gene="VCA0654"
	1854. .2846
	/gene="VCA0654"
	/note="similar to GB:L14685 PID:289864; identified by sequence similarity; putative"
	/codon_start=1
	/transl_table=11
	/product="sucrose operon repressor ScrR, putative"
	/protein_id="AAF96555.1"
	/db_xref="GI:9658070"
	/translation="MASLHDVARLAGVSKSTVSRVINDEYGVYKSTKIKVLKAVRQCGYMQVAKDLKSQKTLNLTIVPRVSSHATQVGLTAIFEQAGKHVLLANTHOVHAKEIYIOLFNOKRVEGIVFYATHLDPLVHALQHSAPVPLVVGQGSVLYNTPISIVHDDHRVGEAGORLVKAGCKQIGFIVQSDDIADVLRSLQGLQALGPHQQTFLPHARGNFSIESGYMAKAYLDIYDLPOLDGLFCATDRLAIVGAIKALQELGVHVQGVQQLVGVNDELAVVNPPLSTFNYPADKAGENAAMKLLERAGRGQEMSKVVLTFQNIQRETCPV"
gene	2922. .4568
CDS	/gene="VCA0655"
	2922. .4568
	/gene="VCA0655"
	/note="similar to GB:M74035 PID:155269; identified by sequence similarity; putative"
	/codon_start=1
	/transl_table=11
	/product="sucrose-6-phosphate dehydrogenase"
	/protein_id="AAF96556.1"
	/db_xref="GI:9658071"
	/translation="MLDLDLLEAGGNNVTRILAPQGVVYALKHPLVPHLPDDVSLOSVLGEWLSQVQRTAEVSDQALAAIGKAIEROKLETLPYQALDCPYRPLWHISPPQGLLNDNGFIYHOGEXHLYFOWHPFACEDKQVWHLKSLDLVDVMOGVALTPSDWFDSHGVSFSGHAYHQDNLWFYTGNTGRVDRQRTQCCARMANQGEFEKLVIRCLPEGVTHIRDPKVIYTGKQWMLLGAQTLAHQGLAVYHSDLLHWHFGLKLYDELGWLDEYDFVAPQTAQADRRVLCGWMGLPDETQHPSCDQGWIHQTLATLELWREGRIYHPLRELDTLOSEPHTLLSDNVTELKTSFALQVTLPMGCELRLMONTQYRVTLTLDENOLLDRSATQIRGKTLRELKDSPTVELRLADOSSLEIFINOGEHVMTSRIFTPRDASGLSHGASVDKLYYMAPASAFENLVNQV"
gene	complement(4690. .5661)
CDS	/gene="VCA0656"
	complement(4690. .5661)
	/gene="VCA0656"
	/note="similar to GB:M76768 SP:P22824 PID:155263; identified by sequence similarity; putative"



JOURNAL	Submitted (24-OCT-2000) Department of Veterinary Pathobiology, University of Minnesota, 1971 Commonwealth Ave., St. Paul, MN 55108, USA		
FEATURES	Location/Qualifiers		
source	1. .10085		
	/organism="Pasteurella multocida"		
	/strain="PM70"		
	/db_xref="taxon:747"		
gene	344..625		
	/gene="PM1841"		
CDS	344..625		
	/gene="PM1841"		
	/codon_start=1		
	/transl_table=11		
	/product="unknown"		
	/protein_id="AAK03925.1"		
	/db_xref="GI:12722267"		
	/translation="MVNCRIGSLSEILHKQLGHQHSIVNCRIGSLKDSRAGRKSYVNC RIGLSLELCIAFFNSTVYVNCRIGSLKQNLHLFHDVYVNCRIGSLERIY"		
gene	1158..1400		
	/gene="PM1842"		
CDS	1158..1400		
	/gene="PM1842"		
	/codon_start=1		
	/transl_table=11		
	/product="unknown"		
	/protein_id="AAK03926.1"		
	/db_xref="GI:12722268"		
	/translation="MQCASFNTRVNCRIGSLSEIFSQDCLLNPVNCRIGSLKQNKNC IAKFIYVNCRIGSLSEMNVRNATFVNCRIGSLEM"		
gene	complement(2259..3224)		
	/gene="PM1843"		
CDS	complement(2259..3224)		
	/gene="PM1843"		
	/codon_start=1		
	/transl_table=11		
	/product="unknown"		
	/protein_id="AAK03927.1"		
	/db_xref="GI:12722269"		
	/translation="MKICLVKTSMDILHSLPALTAQRALPNLQVDVVEESFAEI PWHRSVAVQIPIALRWKPTPTALTARKWKYRTLQSEQYDAVIDAQGLKLSALF ATRLARGTRIGYDRHSREPIAFCYDKKYAISYQQHAVERIQLFAQSLNPLPPTQ GDYIARHPQHAQTMPEYVLFHCTTRQDKFLPDORWRTLAELTAQCIQITHVPSNS QEKORAEIRAGGLDLVTLPKSTLSLAQOIVNASAVVSDTGLAHLASAALDKTNLT YGATDPKLIGSGKNQHYLVNESMDKTEARDEIFQKLTALLSLNTI"		
gene	complement(3221..4267)		
	/gene="rfaf"		
	/note="PM1844"		
CDS	complement(3221..4267)		
	/gene="rfaf"		
	/codon_start=1		
	/transl_table=11		
	/product="Rfaf"		
	/protein_id="AAK03928.1"		
	/db_xref="GI:12722270"		
	/translation="MHILIGPSVGDMMSHSYQTLKQHYPLCEIDVMAPNCKPL LERMPVQAITMPLHGAFALGDRLRGKALRHQYDAIVLPLNSLSAFTPFPAKIP LRWGKGESRYFLNDLRNKKDYPMVQRYIALAFEGKAVPKAEDLP RPVPYLTVDP VTQOHTLKFEKQTLISPRSIGFCFAEFGPAKRWPHYHYAKLAEMLIQQGYAIRL FGSAKDEAVGEIRQAALPETMRHFCINLAGQTSLEAVDLIADCTAIVTNDGLMHTA AAVORPLVALYGPSTPYTPPLSEKAVIIRIEGLIKVRKGDSDGYHQSLIDIQ ELVPDKLTGLLANN"		
gene	complement(4333..4743)		
	/gene="PM1845"		
CDS	complement(4333..4743)		
	/gene="PM1845"		
	/codon_start=1		
	/transl_table=11		
	/product="unknown"		
	/protein_id="AAK03929.1"		
	/db_xref="GI:12722271"		
	/translation="MKSILTOLLFLTLTALLPTSLFAHIVHQAMFSAKAGEPSAIFWN IHNTGNEDVSIALVQSDIRADIVLHGTHQGMIEVTGINLPAHQMTALKRGGHLIMVF		

BASE COUNT  
ORIGIN

2923 a 1988 c 2212 g 2962 t

NVEODLNVDGSPFLRLLLFDNGEIMHIEAKIIQY"  
complement(4859..6283)  
/gene="ptsB"  
/note="PM1846"  
complement(4859..6283)  
/gene="ptsB"  
/codon\_start=1  
/transl\_table=11  
/product="ptsB"  
/protein\_id="AAK03930.1"  
/db\_xref="GI:12722272"  
/translation="MNFQIAQQVIDKLGGKDNISAAACATRLRIVVKNENLIDKKG  
IENIEGVKGQFVAGQYQIIFGSGTVNKVYAALSKLIGIGDMTSEVAAAGTEKQGLL  
ORLVKGLADIEVPILPAIVAGSLMG IHSMLTAGFFVEKNVVDLYPAIADLVDFIN  
TIANAPFVLVLGFSNTRKFGNPFGLAALMLLHPALSDGNNALTLAKNQIY  
WHIFGLEIRVGYQGTIVPLVASWVLATLENKUKVYPSFDNLITPLFLAFTIGLL  
AFTVIGPIREGAGSLISLTGLTWLDTLTFVGGAIFTGLYAPIVITGMHQTIFAETQPL  
LAEVNRTGYRPFITPAAMSNIAOGAAGAAVYMKDAKVRGIAVPSGSIALLGTEPA  
MFGVNLRYRYPFISAMIGAGISSAVIALFNVKATALGAALGPGIPSIKPDLSLAMYCVG  
MLISASTAFTLTVLGKRAOLKAE"  
complement(6512..7531)  
/gene="scrR"  
/note="PM1847"  
complement(6512..7531)  
/gene="scrR"  
/codon\_start=1  
/transl\_table=11  
/product="scrR"  
/protein\_id="AAK03931.1"  
/db\_xref="GI:12722273"  
/translation="MSITNKKRITLSDIACCGSLTTVSMILNNRGEEFKIKAETCQ  
RVKATAEQGYRANVYAKSLKORSNVIGFVDPDLYNGFACTAKILEKOCRENGLOL  
IIACSDNNPQOEKLAIERLLDROIDFLITAPHODPSYQKIGQHTPIQLDRYVPL  
ALNYVISNDSEKIALICRLCDYQLTETFFYGGORTLSPESRLAGFLQGLKQANLN  
LKEMILHRTYQEPESGYMAETAKGLRPLQAIFTASYTLLEGVLYLTQKKMDLL  
LNROLHLATFDHDLNALPFIHSIKQNHPLAQHIFQHMVRFAGKSPNQKIDCD  
VIYRP"  
complement(7557..9056)  
/gene="scrB"  
/note="PM1848"  
complement(7557..9056)  
/gene="scrB"  
/codon\_start=1  
/transl\_table=11  
/product="scrB"  
/protein\_id="AAK03932.1"  
/db\_xref="GI:12722274"  
/translation="MPFSLSKVRSPFCHFTWSPMYIFNQGKYKSLHAQOVGELAQIRQ  
HVLRDTPRPTVHLAPETGLLNDPNGLVFDGEKYHI FYOWPPYAAIHGMKKHMKHPTV  
DFHFQVADPLIPDELFEHSGCYSGALLWQDQIVAFYTGWTRNAENRIPHONTAIF  
KKGDTLLKRCLLNQAPTGYTEHVHDPKAPFITEBGRIFILGAQREDSGLTALYEMT  
DLDATPQLIGBEFLKLPDFDQNVFWMCEPDLLKQDLFLFMSPOGKAQKTYCFQNH  
ATYALGOYANLHFSAHQHIEELDYDFDFAQSIQSGEIGIYAWVGLPDLTYPTDEYQW  
HMSLSLPROFSIQOGLIYORPIPOVYAOCLPAHKYVVEGKQIIPHLDRSYVAFDAENQ  
PFTLREFTNORGQOILLRYONGILTLDRYTTOTALMQQFGEORHCILPOLNHHVEIFF  
DRSVVEIFLNGEKWMTSRFEIADRENALDSRPLCVSAELKPIILTP"  
complement(9037..9963)  
/gene="PM1849"  
complement(9037..9963)  
/gene="PM1849"  
/codon\_start=1  
/transl\_table=11  
/product="unknown"  
/protein\_id="AAK03933.1"  
/db\_xref="GI:12722275"  
/translation="MGKIWLGDVAVDVLIPGENTYLRCAAGAPANVAVGIARLGATT  
GFIGRGDDPLGREMLANKLOQENVCTQVHMLDPKORTSTVLDLKEGERSFTFWNPS  
ADQFLEUSDLPHPNPNMWHCCSIALINEPSRSSTFAAIIHRIKAGGFFSFDPNLRES  
LWNPAEMKTVYNVALADYVKFSEBELFLDSDTSLDATAKLTQHPDKLIIVTL  
KLGATYTLQGSQSVSKKALIPVDTTGTAGDAFVSGLLFGISQVKNQNPNTALVEII  
KQANACALATTAKGMSALPNKQOLDAFLAE"





```
DTEPGSLAAMIDLTAKAEPGGLPFAISDSEAMERKRIKKTKTCTCYCGVCSFEVW
TKDRILTKVQSHSDPANKIVTCVKGFSGHINSQDRLTKPLVRKNGEFHEVEWDEA
LNVIAADNTSGEIKGPDALSPFISSKATNESYLMOKLAROVICNTNVNDCSRVCOA
PATKGLFTVCHGSGSSTEDLEKXAMSVLIGTNTAEAPVIAARMKRAKLFQCKIH
VFDIRKMAERADRFQYKPGQTDLAWSAVTKYLIIDHDLKAFIDEDWDDFDEIYK
SLETFMAFEAEATGIPESLTKFAEACAKAESVVICWAMGITQODIGSDSSTAISNL
LLVTGNYRPPGTGAYPLRGHNNVQCSMDMPKITGTYSQSTEADIRAKFEKEYGVK
LNPAGKNHEVGEIHGDEHSLYLGEDTGVDSNTINFTVQAAFEKIDFMVQDEFL
TFTATYADVLPASPSEKDGFTFTERRIORLYOALEPLGDSKPDWKIFQAIANRLG
EDWAKKHSEINDEVARLITPLIYAGVSYDRLEGFNSLQWPVQDGDDEPILYLEGFNEF
FNKAKLPLSPDNFYQDETIYIHVNNGRLLHFPHEGNTYOTPMIKYKVPRAFVEIS
PELAEDRGIHEGAAYKLISETGEAVLVHVDTRVKGEIYIPLNNDAMENGDLGAINL
LTNSVDQYDTTPSYKRTSCRLEVTIKRGKSLPNPNFRVNNKRHPQYSVQVQKKWER
SDYVFPGNQVDK"
complement(7804..8751)
/gene="SAV2310"
CDS
complement(7804..8751)
/gene="SAV2310"
/Note="similar to lyt divergon expression attenuator LytR"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB58472.1"
/db_xref="GI:14248084"
/translation="MKRSKSKMSLPKFIWVFGILVLAIVAVVYVAAKIFITGNK
IHNPLDRHSELRAEIVSGDPTFIALFGVDSADRRKKGGGERSDSIMLSINPK
TKKTEIVSPDRTRAEIVRGRTGTEKIAHAYAGGPNMAKSKLEKLMNVPIIDHYATIDM
DGLHNMIDSIGDVVDVNDFTVDGVRETGKQOOTHVNGDOALKFIRSRKEGAGDGF
RQORQIVLEAMKAIASPSITHENSMLNEIQNVKTDLGLDLNTIRSNYKANDT
INKHLSQGGGQISDGLYFIPSEQSAESTKLLKNDLE"
complement(8884..9681)
/gene="SAV2311"
CDS
complement(8884..9681)
/gene="SAV2311"
/Note="similar to suppressor protein subB"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB58473.1"
/db_xref="GI:14248085"
/translation="MDKTLQIDKILCSWLKQIDNVIPOLIMETTTETKRHRFLVLT
NVDKIQOQOFLATHPEHQLLAESNEMITNEIHLNIMDPIDGTANLVKQOED
YCIILAYEYEGKPLSYVDYVPHKKLYKAIIRGEAFNGKIRMEPPSLKLEDAIISFN
AQVMNLDTVQDLFDASFSYRLVAGCLDSMRVAKQFGAHINTNPKWDIAAQFLFAE
LLNKMWTLDGKAIDHLKGAFLIISKACHETVLKILNANGYQYR"
10066..10758
/gene="SAV2312"
CDS
10066..10758
/gene="SAV2312"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB58474.1"
/db_xref="GI:14248086"
/translation="MKAERONLIITAIQONKMTALELAKYCNVSKRTILRIDDL
NOGVKIYAHGCKNGCYOIQOAGKIALNSETQLSALFLVNESOSYSTLPVKSEINA
IILKQCLSPQTRLRLLKRMDEYIKEDTQHTMLPMLFSDILLYCTERNVMLVDRVD
DNIAKQVIFGLLCKHGHWAVIDYIAQDKTALELENIIDISYFQKITQTRDISI
DNYHQLFNPIDS"
complement(10783..11520)
/gene="SAV2313"
CDS
complement(10783..11520)
/gene="SAV2313"
/Note="similar to protein of pX02-46"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB58475.1"
/db_xref="GI:14248087"
/translation="MENEKKYTFKDIARWDLIIPIIGVIFYALTIAAYIGPILFGY
YEIDNLIFFATITQALAYTIGIGAFYFLHKLMPSRIRAGFAYIKKWLRLSIYII
ALILIITYEYKQFQKLOYSFANELELNKMFVPAFLPVAFLLIIVGVPIVEEIV
FRHILIGELGKKNFIAMSVSVFLFAFIHVTDAKSPFEFGYLLILSVLYLKSG
```

```

RNLGSTIALHIANNEVSVFISVISIYA"
11569..11898
/gene="SAV2314"
CDS
11569..11898
/gene="SAV2314"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB58476.1"
/db_xref="GI:14248088"
/translation="MINETKMRQOITNLIQIILDIKYNLTMDNIOHTHIYKFNQIV
TNELNKFQFFETSYIINENQVSYEIIILLNKRPLRQVDYGNKIQYLNIFYHTLSNAL
FAIKFAH"
12343..13215
/gene="SAV2315"

Query Match      10.9%; Score 121; DB 1; Length 341350;
Best Local Similarity 55.9%; Pred. No. 2.1e-22;
Matches 279; Conservative 0; Mismatches 205; Indels 15; Gaps 2;

QY 1 tatgatttcggcggtccagtcggcggtctgctctcggtctggtctactcaccacatcgtc 60
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 76239 TATGAATTTGGTGGTCAATTTGGTGAATTAATATTCGGATTATTGTATGTCGCGATTGTT 76180

QY 61 atcactggtctgcaccagtccttcccgccaattgagctggagctgtt-----t 108
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 76179 ATTACAGGTATGCATCATAGCTTTATAGCTGTAGAAACGACATTAATTGCAGATGCCACT 76120

QY 109 aaccagggtggatcccttcattcttcgcaacggcatctatggcctaataatcgccccagggtg 168
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 76119 AAAACGGTGGATCATTTATATTCGCGATTGCGACAATGCTCTAATGTTGCACAAGTGGT 76060

QY 169 gcatgtttggcagtgcttcttcctgcgcga---agagtgaagaaagctcaaggcccttgcaggt 225
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 76059 GCAGCAATTCGACGGTCTTTATTATTAACAAAAATAAGAGTTAAAGGTGTGGCATCT 76000

QY 226 gcttcaggtgtctccgctgtcttggtattacggagcgtcgatcttcggtgtaaacctt 285
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 75999 GCGCGAGGTATTTACAGCATTTACTTGTGTAATACAGAACCGGCTATGTTGGTGTAACTTA 75940

QY 286 cgcctgcgctg9ccgttcttcacatcggtatcggtaccgcagcgtatcggtg9ccgtttgatt 345
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 75939 AAACATAAGATATCCATTTATTTGGCGCTATCGTTGGATCAGGTATTGTTTCAGCATATATT 75880

QY 346 gcactctttaataatcaagcagtcggttcggttcggttcggttcggttcggttcggttcggt 405
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 75879 GCTTTCTTCAAGTTAAAGCAATCGCATTAGGAACCTGCGATTGCCAGGATTTATTTC 75820

QY 406 attgatgtcccagatagtggtcatgttcttggtgtgtgagcagttgttaccttcttcacgcga 465
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 75819 ATCAATCCAGTACATGCAGGATGGTTACACTACTCTTTGTTGGTATGACATATCATTCATC 75760

QY 466 ttcggcgccagcagtgctt 484
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 75759 ATTGCTATAACAGTACTT 75741
```

Search completed: March 22, 2002, 08:55:40  
Job time: 8049 sec



Result No.	Score	Query %		Length	DB	ID	Description
		Match					
1	666	36.6	650	2	Q9S6S6		lactococcus
2	562.5	30.9	630	2	Q9KG19		Q9S6S6 bacillus ha
3	550.5	30.2	636	2	Q9KFP0		Q9KG19 bacillus ha
4	539.5	29.6	674	2	Q9S6S6		Q9KFP0 streptococc
5	532.5	29.2	627	2	Q9LHG6		Q9S6S6 clostridium
6	517	28.4	644	2	Q9KJ80		Q9LHG6 streptococc
7	494	27.1	620	2	Q9S6S6		Q9KJ80 streptococc
8	480.5	26.4	616	2	Q9S6S6		Q9S6S6 clostridium
9	479	26.3	577	2	Q9L461		Q9S6S6 lactobacill
10	468	25.7	621	2	Q9K408		Q9L461 klebsiella
11	448	24.6	636	2	Q9CFK9		Q9K408 lactococcus
12	445.5	24.5	620	2	Q9A0X4		Q9CFK9 streptococc
13	433.5	23.8	618	2	Q9S6S6		Q9A0X4 enterococcu
14	418.5	23.0	480	2	Q9S6S6		Q9S6S6 staphylococ
15	416	22.8	479	2	Q9KLT8		Q9S6S6 vibrio chol
16	411	22.6	661	2	Q9K408		Q9KLT8 corynebacte
17	398.5	21.9	640	2	Q9RLJ2		Q9K408 listeria mo
18	390.5	21.4	683	2	Q9K408		Q9RLJ2 corynebacte
19	350.5	19.2	470	2	Q9KAS1		Q9K408 bacillus ha

Db	342	GVMLRTNKKL	KALAGASTS	ITALFGTTEP	AVGYVTLP	LKKPFTM	AVISAAVGGAI	VGHYG	401
Qy	120	IKAVALGAAG	FLGVVSI	--DAPDMV	MFLCAVV	VTFFIA	FGAAIAI	YGLXIVLRNGSIDP	176
Db	402	SVAVAPGAP	GLLTIP	IFY	PEDGRG	FVAFV	IAIIISFLA	----AVLTIVGFKDPVDD	456
Qy	177	ATAAPV	PAGTTK	KABAPAF	SNDSTII	IQAPL	TGAEIAI	SSVSAMDAMPASGKLGSGVAIYP	236
Db	457	DTLSN	ESGSENV	EKREDDK	EP	SASEBI	EIKSPL	KGVEVPLTEVODHVFFSSGAMCKGVAVRP	516
Qy	237	TGOLV	SPVSGKI	VVAPPS	GHAFAVR	T	KAEDG	SNVDIMHIGFPTNLNGTHFNPLKKOG	296
Db	517	KEGLV	API	NGTV	TSLE	T	KHAIGTS	----DNGTEIF	572
Qy	297	DEVKAG	ELLCE	FFDIDA	I	KAAGY	EVTTPT	IVVSNYKKTGPVNTYGLGIEIAGANILLNAK	354
Db	573	DEVAAG	DLLEF	DVER	ITAAGY	DVIT	PLVLT	INAKOFSNVOTDKREVTSEDLIIHVIK	630

DE PTS SYSTEM, BETA-GLUCOSIDE-SPECIFIC ENZYME II, ABC COMPONENT;  
DT 01-JUN-2001 (TREMBler. 17, Last annotation update)

RC STRAIN-C-125 / JCM 9153;  
RX MEDLINE=20512582; PubMed=11058132;  
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,  
RA Fuji F., Hirana C., Nakamura Y., Ogasawara N., Kuhara S.,

RA	HoriKoshi K.;
RT	"Complete genome sequence of the alkaliphilic bacterium <i>Bacillus</i>
RT	<i>halodurans</i> and genomic sequence comparison with <i>Bacillus subtilis</i> ."
RL	Nucleic Acids Res. 28:4317-4331(2000).
DR	ENBL; AP001509; BAB04314.1; "
DR	InterPro; IPR001127; PTS_EIIA.
DR	InterPro; IPR001996; PTS_EIIB.
DR	InterPro; IPR003352; PTS_EIIC.
DR	Pfam; PF00358; PTS_EIIA_1.1.
DR	Pfam; PF00367; PTS_EIIB.1.
DR	Pfam; PF02378; PTS_EIIC.1.
DR	ProDom; PD001476; PTS_EIIB.1.
DR	ProDom; PD002243; PTS_EIIA.1.
DR	PROSITE; PS00371; PTS_EIIA_1.1.
DR	PROSITE; PS01035; PTS_EIIB_CYS.1.
KW	Complete proteome.

```

DR ProDom; PD001476; PTS_EIIB; 1.
DR DR Dom; PD002243; PTS_EIIA; 1.
DR DR PROSITE; PS00371; PTS_EIIA_1; 1.
DR DR PROSITE; PS01035; PTS_EIIB_CYS; 1.
KW Complete proteome.
SQ SEQUENCE 636 AA; 68437 MW; 434C0B12311716F2 CRC64;

```

Db 294 YNLSPIIAGAEFLGGFWOVFVIFGLHWGLIPIAINNLVVOGSDPVLAMVFAASFAOIGAVA 353

Db 354 AVWLKIQKQVKVLSVPAPFISGIFGVTPEAIVGYVTLPLKPTFIISCAAAVGAAGIUGFR 413

QY 120 IKAVALGAAGFLGWSEI---DAPDMVMF-LVCAVVTFTAFCAAIYAGLYLVRRNGSID 174

Db 414 SQGYIIGGLGIFGIPSFLLPADGMDAGFWGIATVAVVAFV-LGFILTYLFGILKSGNASDE 472



```

OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NG8;
RX MEDLINE=20340959; PubMed=10878120;
RA Cote C.K., Cvitkovich D., Blewels A.S., Honeyman A.L.;
RT "A novel beta-glucoside-specific PTS locus from Streptococcus mutans
RL Microbiology 146:1555-1563(2000).
DR EMBL: AF206272; AAF89975.1; -.
DR InterPro: IPR001127; PTS_EI1A;
DR InterPro: IPR001996; PTS_EI1B;
DR InterPro: IPR003352; PTS_EI1C;
DR Pfam: PF00367; PTS_EI1A_1; 1.
DR Pfam: PF02378; PTS_EI1B; 1.
DR Pfam: PF02378; PTS_EI1C; 1.
DR ProDom: PD002243; PTS_EI1A; 1.
DR PROSITE: PS00371; PTS_EI1A_1; UNKNOWN_1.
SQ SEQUENCE 644 AA; 69282 MW; 723B7FBBDD2794EB CRC64;

Query Match 28.4%; Score 517; DB 2; Length 644;
Best Local Similarity 35.0%; Pred. No. 5.5e-24;
Matches 126; Conservative 65; Mismatches 149; Indels 20; Gaps 8;

QY 1 YDFGGVGLLGLVSPVITGLHQSFPPIE--LELFNOGGSFIFATASMANIAQGAACL 59
DB 299 YGFNPVYIGVLCAMQVLMFGLHGLVPLAILELQKGGVILVATIAIC-FAQAGSLL 357

QY 60 AVFLAKSEKLKGLAGAGSVAVLGTETPAIFGVNLRWRPFFIGTGAAGGALIALEN 119
DB 358 NIMRTNENKVRQLSPAFISALFGVTEPAIYGITLPMRVPTMTCSGASGAYLALFN 417

QY 120 IKAVALGAAGFLGVSDIDAPD----MMELVCAVVTFFTAAGAAIAYGL-YLVRNGSID 174
DB 418 VKMQVGGMGLFAIPSDIPDKNSMLIHFLIAIAMPFVLGVLTQFIKIPYLGEPTSTD 477

QY 175 PDATAAPVPAGTTKAEAPAEFSDNSTIIQAPLTGEAIALSSVDSAMFASGLSGSVAI 234
DB 478 SDVDDKEEPVKELK-----EIKQE--IISPLICKVVKLENVPDEVFASGAMKGIAI 528

QY 235 VPTKGLVSPVSKIVVAPSPGHAFVTRKAEDGSNVYDILMHIGFTVNLNGTHFNPLKK 294
DB 529 DPDDGIVAPTKGEVTLVFTKHAVALRT--ENGA--EILIHGMDTVSLAGKGFKSFE 584

QY 295 QGDEVKAGELLCFEDAIKAAGYEYVTPPIVSNYKKTGPVNTYGLGETEAGANLLNVAK 354
DB 585 VGDHVEAGOTLLEFDVNAIKAGLPVITPVIVTNSQDFEDVLTQERTVEAGNYLLTAVK 644

RESULT 7
ID Q99Y91 PRELIMINARY; PRT; 620 AA.
AC Q99Y91;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DE PUTATIVE SUCROSE-SPECIFIC PTS PERMEASE, ENZYME II.
GN SCRA OR SPV1815.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SF370;
RX MEDLINE=21192684; PubMed=11296296;
RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
RA Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,
RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,
RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).

DR EMBL: AE006608; AAK34540.1; -.
KW Complete proteome.
SQ SEQUENCE 620 AA; 65725 MW; 64F5FE83524DC4DC CRC64;

Query Match 27.1%; Score 494; DB 2; Length 620;
Best Local Similarity 36.8%; Pred. No. 1.4e-22;
Matches 126; Conservative 47; Mismatches 149; Indels 20; Gaps 5;

QY 3 FGGPVGGLLGLVSPVITGLHQSFPPIE--LELFNOGGSFIFATASMANIAQGAACLAV 61
DB 295 FG--IAGLIVGGIQQLIIVVTGIIHIFNLEAQLIANTKDPFNAYLTATAAAGATLAV 352

QY 62 FFLAKSEKLKGLAGAGSVAVLGTETPAIFGVNLRWRPFFIGTGAAGGALIALEN 121
DB 353 AVTKSTKTLKGLAFPTLSALLGITEPAIFGVNLRPKVFSGLIGALGGWAGLFGIA 412

QY 122 AVALGAAGFLGVVSDIDAPDMVMELVCAVVTFFTAAGAAIAYGLYLVRNGSIDPDATAAP 181
DB 413 GTGFGITVLPGLTLLYLNGLQLQVLTMLVGLGVAFAIAYTWGY-----QDRETL 462

QY 182 VPAGTTKAEAPAEFSDNSTIIQAPLTGEAIALSSVDSAMFASGLSGVAIVPTKGQL 241
DB 463 LPAVEVDQADQAPALAE--TLYSPLNGTVVDSLAVSDPVFSSGAMGQGLAIKPEDNTL 519

QY 242 VSPVSGKIVVAPSPGHAFVTRKAEDGSNVYDILMHIGFTVNLNGTHFNPLKKQGVDEKA 301
DB 520 YSPVSGKIVVFTGHAYITS-----SQGAENVLLHIGIDTESMAGDGFESLVAVGQAVK 575

QY 302 GELLCEFDIDAKAAGYEYVTPPIVSNYKKTGPVNTYGLGEI 343
DB 576 GDLGHGFDPSKIAEAGLDDTMMIVSNIADYQSDVILAQGHV 617

RESULT 8
ID Q46129 PRELIMINARY; PRT; 616 AA.
AC Q46129;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE PTS-DEPENDENT ENZYME II.
GN ABGF.
OS Clostridium longisporum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1523;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B6405;
RX MEDLINE=98151780; PubMed=9491080;
RA Brown G.D., Thomson J.A.;
RT "Isolation and characterisation of an aryl-beta-D-glucoside uptake and
RL species C. longisporum.";
RL Mol. Gen. Genet. 257:213-218(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=B6405;
RA Brown G.D., Thomson J.A.;
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL: L49336; AAC05713.1; -.
DR HSP: P20166; IGPR.
DR InterPro: IPR001127; PTS_EI1A;
DR InterPro: IPR001996; PTS_EI1B;
DR InterPro: IPR003352; PTS_EI1C;
DR Pfam: PF00367; PTS_EI1A_1; 1.
DR Pfam: PF00367; PTS_EI1B; 1.
DR Pfam: PF02378; PTS_EI1C; 1.
DR ProDom: PD001476; PTS_EI1B; 1.
DR ProDom: PD002243; PTS_EI1A; 1.
DR PROSITE: PS00371; PTS_EI1A_1; UNKNOWN_1.
DR PROSITE: PS01035; PTS_EI1B_CYS; 1.
```

```
SQ SEQUENCE 616 AA; 65890 MW; 7D9CDB1B17BE9283E CRC64;

Query Match 26.4%; Score 480.5; DB 2; Length 616;
Best Local Similarity 31.9%; Pred. No. 9e-22;
Matches 115; Conservative 62; Mismatches 151; Indels 33; Gaps 6;

Qy 1 YDFGGPVG-LGLFGLVSPVITGLHOSFPPIEL-ELFNQGGSFIFATASMANIAOGAAC 59
Db 278 YNFNTIAGLFIQGGVQVFMFGLHWGLVPIAMNNAVIGDPLATVAVCFAGTVVM 337
Qy 60 AVFFLAKSEKLGKLAGAGSVAVLGIPEPAIFGVNLRWPFFFIGTAAIGGALIALFN 119
Db 338 ATLAKTKOKKLKLCIPALISGFFGVTEPAIVGITLPRKKPFILSCIAAGVTGGIIGFFE 397
Qy 120 IKAVLAGAAGFLGVVSDAPD-----MVFLVCAVVFIFAFGAIAIYGLVLRNG 171
Db 398 SKGYSMGGLGIFALPSYINPEGIDRGFGVMVAVIGVGFIL-----MFTVKLND 449
Qy 172 STDPTAAPVPAGTTKAEAEAPAEFSNDSTIIQAPLTGEATIALSSVDAMFASCKLGSG 231
Db 450 EEEVKTTSKEESLVKQEE-----IVSPIOGEVVTLEAVKDEAFSSGALGKG 497
Qy 232 VAIPTKGQVSPVSGKIWVAFPPSGHAFVRTKAEDGNSVDILMHIGFDTVNLNGTHFNP 291
Db 498 VAINPIEGKVYAPADGTLTLFPLSHALGIT--ENGA--ELIIHGMVTVLECKHETA 553
Qy 292 LKQDEVKAGELCEFDIDAIAKAGYEVYTPIVVSNYKKTGPNVNTYGLGETEAGANLNL 351
Db 554 KVKQDGKIKKQGLLEFDKEAKYAGYTIPTVLITNSDYLDVTDTKRKVDVNSSELLT 613
Qy 352 V 352
Db 614 V 614

RESULT 9
Q9LA61 PRELIMINARY; PRT; 577 AA.
ID Q9LA61
AC Q9LA61;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE PERMEASE.
GN BGIP.
OS Lactobacillus plantarum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
OC Lactobacillus.
OX NCBI_TaxID=1590;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B21;
RX MEDLINE=20263773; PubMed=10802183;
RA Marasco R., Salatiello I., De Felice M., Sacco M.;
RT "A physical and functional analysis of the newly identified bgIPCT
operon of Lactobacillus plantarum."
RL FEMS Microbiol. Lett. 186:269-273(2000).
DR EMBL; AJ250202; CAB71150.1;
DR InterPro; IPR001127; PTS_EIIA.
DR InterPro; IPR001996; PTS_EIIB.
DR InterPro; IPR003352; PTS_EIIC.
DR Pfam; PF00358; PTS_EIIA.1;
DR Pfam; PF00367; PTS_EIIB.1;
DR Pfam; PF02378; PTS_EIIC.1;
DR ProDom; PD002243; PTS_EIIA.1;
DR PROSITE; PS00371; PTS_EIIA.1;
DR PROSITE; PS00371; PTS_EIIB.CYS.1;
SQ SEQUENCE 577 AA; 60877 MW; E8C877050103B5FC CRC64;

Query Match 26.3%; Score 479; DB 2; Length 577;
Best Local Similarity 34.1%; Pred. No. 1e-21;
Matches 118; Conservative 58; Mismatches 88; Indels 82; Gaps 11;

SQ SEQUENCE 621 AA; 65492 MW; DFIDE2A2A7B81339 CRC64;

Query Match 25.7%; Score 468; DB 2; Length 621;
Best Local Similarity 33.0%; Pred. No. 5.3e-21;
Matches 120; Conservative 67; Mismatches 129; Indels 48; Gaps 9;

Qy 7 VGGLLFGLVSPVITGLHOSFPPIEL-ELFNQGGSFIFATASMANIAOGAACLAFFFLA 65
Db 290 LAGAAAGALWQVCVIFGLHWGLIPLMINNAVILGHDSMMPMLLPVAMGVGALGIFLRT 349
Qy 66 KSEKLGKLAGAGSVAVLGIPEPAIFGVNLRWPFFFIGTAAIGGALIALFNKAVAL 125
```

```
Db 350 RDARQKVLGASAVSAGIFGVTEPAIYGLNPLRRPFIFGCVYWPIGAMVG-FSDSHVST 408
QY 126 GAAGFLGVVS-----IDAPDMWFLVCVVTFIAFGAAIAYGLYLVRNGSDTDP 176
Db 409 YSFGFNIITLAQMIPEGIDAT-----VWGGAAGMFASLII----- 445
QY 177 ATAAPVPAGTTKAAEAE-----APAEFSNDSTIIQAPLTGEAIALSSVSDFAMFASGLKGS 230
Db 446 ACVLTIVAGLPRSSAEQAQVWAPASV-ND---ILAPMTGSVLALDQVPDSTFASGLLGQ 501
QY 231 GVAIVPTKGLVSPVSGKIVVAFPSGHAFVTRKAEDGSNVDILMHIGFTVNLNGTHFN 290
Db 502 GVAIISVGVKVIAPFSGVASIPQTHAIGLIS-----DSGIELLIHVGDITVKLDGAPET 557
QY 291 PLKKQGVDEKAGELLCEFFIDAIKAAGYEVTTPIVVSNYKKTGPVNTYGLGETEAGANLL 350
Db 558 AHVKEGDKIKAGDULLEFFDRLQALDAGYDLATPIIISNSDDFTLDVWSASAVDAQPLL 617
QY 351 NVAK 354
Db 618 SVSR 621

RESULT 11
Q9CFK9
ID Q9CFK9 PRELIMINARY; PRT; 636 AA.
AC Q9CFK9;
DT 01-JUN-2001 (TREMREL. 17, Created)
DT 01-JUN-2001 (TREMREL. 17, Last sequence update)
DT 01-JUN-2001 (TREMREL. 17, Last annotation update)
DE BETA-GLUCOSIDE-SPECIFIC PTS SYSTEM IIABC COMPONENT (EC 2.7.1.69).
GN PTBA.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Lactococcus.
OX NCBI_TaxID=1360;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TL1403;
RA Bolotin A., Wincker P., Mauger S., Jaillon O., Malarne K.,
RA Weissenbach J., Ehrlich S.D., Sorokin A.;
RT "The complete genome sequence of the lactic acid bacterium Lactococcus
RT lactis."
RL Genome Res. 0:0-0(2001).
DR EMBL; AB0063376; AAK05558.1; -.
DR InterPro: IPR003439; ABC transport.
DR InterPro: IPR001127; PTS_EIIB.
DR InterPro: IPR001996; PTS_EIIB.
DR pfam: PF00358; PTS_EIIB_1; 1.
DR pfam: PF00367; PTS_EIIB; 1.
DR pfam: PF02378; PTS_EIIC; 1.
DR ProDom: PD001476; PTS_EIIB; 1.
DR ProDom: PD002243; PTS_EIIB; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 636 AA; 67630 MW; 012030D819163325 CRC64;
```

```
Query Match 24.6%; Score 448; DB 2; Length 636;
Best Local Similarity 33.0%; Pred. No. 9.2e-20;
Matches 110; Conservative 59; Mismatches 136; Indels 28; Gaps 7;
```

```
QY 3 FGGPVGLLFLGLVSPVITGLHOSFPPIELELFNQG-CSFIFATASMANIAQGAACLA 61
Db 288 FNPIIFGLVGFQWQVLMFGLHWAIVPFAIILAKGEPTALLIAASVASFAQTGA 347
QY 62 FFLAKSEKLGAGSGVSAVLGITEPAIFGVNLRWPFIFIGTAAIGGALIALFNK 121
Db 348 MLKTKDKRLRELAIPAFISGWFGVTEPAIYGITLPKKRPFVWSCIVSGILSA 407
QY 122 AVAALGAGFLGVISDAPDMWFLVCVVTFIAFGAAIAYGL---YLVRNGSDTDP 178
```

```
Db 408 AYTMA-----LGIFSFTSNTILTGEVSGAIKIMIVSAVAVIAGFVVTYLV---GFEDDVI 460
QY 179 AAPVPAGTTKAAEAEPAEFSNDST---IIQAPLTGEAIALSSVSDFAMFASGLKGS 235
Db 461 ENPIP-----DKFNKQKTNEIIGSPLEGKVIPLSQVKDAFSAVWGKAAIE 510
QY 236 PTKGQLVSPVSGKIVVAFPSGHAFVTRKAEDGSNVDILMHIGFTVNLNGTHFN 295
Db 511 PTLGEVRAFPDGMVMILFPTKHAVGLIS-----NEGTELLIHIGIDTVQLGKGFET 566
QY 296 GDEVKAGELLCEFDIDAIKAAGYEVTTPIVVSN 328
Db 567 GOSVKKGDLLKFDIERIQIONAGYSTQVPIVTN 599

RESULT 12
Q9A0X4
ID Q9A0X4 PRELIMINARY; PRT; 620 AA.
AC Q9A0X4;
DT 01-JUN-2001 (TREMREL. 17, Created)
DT 01-JUN-2001 (TREMREL. 17, Last sequence update)
DT 01-JUN-2001 (TREMREL. 17, Last annotation update)
DE BETA-GLUCOSIDE PERMEASE IIABC COMPONENT.
GN SPY0572.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SF370;
RX MEDLINE=21192684; PubMed=11296296;
RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
RA Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,
RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,
RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes."
RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
DR EMBL; AE006513; AAK33557.1; -.
KW Complete proteome.
SQ SEQUENCE 620 AA; 66148 MW; 1AF6872CFDD7C7D6 CRC64;
```

```
Query Match 24.5%; Score 445.5; DB 2; Length 620;
Best Local Similarity 33.0%; Pred. No. 1.3e-19;
Matches 107; Conservative 63; Mismatches 131; Indels 23; Gaps 7;
```

```
QY 11 LFLGVYSPVITGLHOSFPPIELELFNQGSFIFATASMA-NIAQGAACLAFLAKSE 69
Db 288 ILGAIMPLVVMVTGMHWAFAPIFLAASVATPDVLILPAMLASNLAQGAASLAVAFKTK 347
QY 70 LKGLAGASGVSAVL-GITEPAIFGVNLRWPFIFIGTAAIGGALIALFNKAVAGAA 128
Db 348 TRQVALAAGISALLAGITEPALYGVTLKFKPLYAAMISGGLVGAFIGFVNIASTY 407
QY 129 FGLGVVSDAP---DMWFLVCVVTFIAFGAAIAYGLYLVRNGSIDPDATAAPVPA 184
Db 408 SIIGLPQYINPSGGANFTNALIAGTATIVLAFSLTFWFG-----IDEE---SPK 455
QY 185 GTTKAAEAPAEFSNDSTIIQAPLTGEAIALSSVSDFAMFASGLKGSVAIVPTKGL 244
Db 456 SVAADMSQVKSGLSTKQT-LYAPMTGEMFLSEVPDTEFSSKLLGEGFALLPGEVYAP 514
QY 245 VSGKIVVAFPSGHAFVTRKAEDGSNVDILMHIGFTVNLNGTHFNPLKKQGVDE 304
Db 515 FDEVEITFPTKHAVALK-----NTRGVEVLHVIGIDTVLKGQGFQQLVSGVDV 570
QY 305 LCEFFIDAIKAAGYEVTTPIVVSN 328
Db 571 LLKMDIDFITSKGYSLISPVVVTN 594

RESULT 13
```

```
Q9X565
ID Q9X565 PRELIMINARY; PRT; 618 AA.
AC Q9X565;
DT 01-NOV-1999 (TReMBLrel. 12, Created)
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE BETA-GLUCOSIDE SPECIFIC TRANSPORT PROTEIN.
GN BGLS.
OS Enterococcus faecium (Streptococcus faecium).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Enterococcaceae;
OC Enterococcus.
OX NCBI_TaxID=1352;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=BFE 900;
RC MEDLINE=99240446; PubMed=10224016;
RA Franz C.M.A.P., Worobo R.W., Quadri L.E.N., Schillinger U.,
RA Holzapfel W.H., Vederas J.C., Stiles M.E.;
RT "Atypical genetic locus associated with constitutive production of
RT enterocin B by enterococcus faecium BFE 900.";
RL Appl. Environ. Microbiol. 65:2170-2178(1999).
DR EMBL; AF121254; AAD28228.1; -.
DR HSSP; P20166; IAX3.
DR InterPro; IPR001127; PTS_EIIB.
DR InterPro; IPR001996; PTS_EIIB.
DR InterPro; IPR003352; PTS_EIIC.
DR Pfam; PF00358; PTS_EIIB_1; 1.
DR Pfam; PF00367; PTS_EIIB; 1.
DR Pfam; PF02378; PTS_EIIC; 1.
DR ProDom; PD002243; PTS_EIIB; 1.
DR PROSITE; PS00371; PTS_EIIB_1; 1.
DR PROSITE; PS01035; PTS_EIIB_CYS; 1.
SQ SEQUENCE 618 AA; 67051 MW; 9AECDC16BF2DEC9C CRC64;

Query Match 23.8%; Score 433.5; DB 2; Length 618;
Best Local Similarity 35.1%; Pred. No. 6.9e-19;
Matches 128; Conservative 62; Mismatches 130; Indels 45; Gaps 15;

QY 7 VGLLFLGLVYSP----IVITGLHQSPFPIELFLNQGG---SFIFATASMANIAOAGACL 59
DB VGSWFVFMVLYSALQPFITMLGAGNFIMPVVASLIAANGYDPAFI--SSCTISDIAYGGAML 336
QY 60 AVFFLAKSEKLGAGASGVSAVLGITPAIFGVNLRWRPFPIGTAAGGALIAFLFN 119
DB GFELTRNLKQQLGFTVLSVLGVTPEPAIVGVVRRPVPVAVMVGGLGGLFAGLTS 396
QY 120 IKAVALGAAGFLGVVS-IDAPDMV--MFLVCAVTFVFFIAFGAAIAYGLYLRNRGSDPD 176
DB VKAYSV-AWGLFGLPAYIGEGDFMFWFVLAADVISA--GSAVA--AYLL-----442
QY 177 ATAAPVPAGTTKAEA--EAPAEFSDSTIIQAPL-----TGEAIALSSVSDAMFASGKLGS 230
DB ----GVPAQEEDEQETINIIAEKGDKNLRTSSLGNVVEGEVISLNEVKDQAFSTGALCK 498
QY 231 GVAIVPTKGLVSPSGKIVAFPSGHAFVTRKAEKDSNVDILMHIGFDTVNLNGTHEN 290
DB GAIKPIDTHLYSPVDGETVVPFNNHIGI--QSDQG--IEILLHIGIDTVLEGGKYN 554
QY 291 PLKKGDEVKAGELCEFDIDAIIKAAGYEVITPIVVSN---YKKTGPVNTYGLGIEAGA 347
DB LNIKGDKIQRCOLLGTVDKIQEAGYDTSTIVVVTWNTSDSLVDVIPSNS---EQVIKTD 611
QY 348 NLLNV 352
DB NLLNV 616

RESULT 14
Q99RQ0
ID Q99RQ0 PRELIMINARY; PRT; 480 AA.
AC Q99RQ0;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
```

```
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE PTS SYSTEM, SUCROSE-SPECIFIC IIBC COMPONENT.
GN SCRA OR SA2167.
OS Staphylococcus aureus subspecies aureus N315.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Staphylococcus.
OX NCBI_TaxID=158879;
RN [1]
RP SEQUENCE FROM N.A.
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.I., Nagai Y., Lian J., Ito T., Kanamori M.,
RA Matsumaru H., Maruyama A., Murakami H., Hosoyama A., Mizutani-Ui Y.,
RA Takahashi N.K., Sawano T., Inoue R.I., Kaito C., Sekimizu K.,
RA Hirakawa H., Kuhara S., Goto S., Yabuzaki J., Kanehisa M.,
RA Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T., Hattori M.,
RA Ogasawara N., Hayashi H., Hiramatsu K.;
RT "Whole genome sequencing of methicillin-resistant Staphylococcus
RT aureus.";
RL Lancet 357:1225-1240(2001).
DR EMBL; AP003137; BAB43469.1; -.
KW Complete proteome.
SQ SEQUENCE 480 AA; 51232 MW; D88607F6E0AF2E84 CRC64;

Query Match 23.0%; Score 418.5; DB 2; Length 480;
Best Local Similarity 49.7%; Pred. No. 4.1e-18;
Matches 89; Conservative 32; Mismatches 47; Indels 11; Gaps 4;

QY 1 YDFGGPVGLGLGLVYSPIVITGLHQSPPIELFLF---NNGGSFIFATASMANIAOQA 56
DB YFEGGAIGLFLGLLYAPVITGMHHSFIAVETTLIADATKGTGGSFIFATMSNVAQGG 360
QY 57 ACLAVFFFLAK-SEKLGLAGASGVSAVLGITPEPAIFGVNLRWRPFPIGTAAGGALI 115
DB AIAIAAFTTIKONKLGKVASAAGISALLGITPEPAMFGVNLKLYPFPIGTAAGGSAVI 420
QY 116 ALFNKKAVALGAAGFLGVVSIDA--PDMVFLVCAVTFVFFIAFGAAIAYGLYLRNRGS 172
DB AFFKVKATALTAGLPGFISINPVHAGWLHVFYGMTISFII---AIVTLLLSKRKAN 475

RESULT 15
Q9KLT8
ID Q9KLT8 PRELIMINARY; PRT; 479 AA.
AC Q9KLT8;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE PTS SYSTEM, SUCROSE-SPECIFIC IIBC COMPONENT.
GN VCA0653.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EL TOR N16961 / SEROTYPE O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Emcleva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae";
RL Nature 406:477-483(2000).
DR EMBL; AE004395; AAF96554.1; -.
DR TIGR; VCA0653; -.
DR InterPro; IPR001996; PTS_EIIB.
DR InterPro; IPR003352; PTS_EIIC.
DR Pfam; PF00367; PTS_EIIB; 1.
```

Search completed: March 21, 2002, 16:28:28  
Job time: 338 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 21, 2002, 16:28:49 ; Search time 17.08 seconds  
(without alignments)  
777.088 Million cell updates/sec

Title: US-09-604-231-4  
Perfect score: 1821  
Sequence: 1 YDFGGPVGLLGLVLPVIV.....IEAGALLNVAKKEAVPATP 362

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	711	39.0	664	1	PTSA_STRMU
2	706.5	38.8	651	1	PTSA_PEDPE
3	515	28.3	609	1	PTBA_BACSU
4	492.5	27.0	631	1	PTBA_ERWCH
5	450	24.7	625	1	PTBA_ECOLI
6	446.5	24.5	480	1	PTSB_STAXY
7	418	23.0	474	1	PTSB_PASMU
8	412	22.6	479	1	PTSB_VIBAL
9	386.5	21.2	674	1	PTGA_CORGL
10	330	18.1	470	1	PTTB_BACSU
11	313	17.2	651	1	PTAA_KLEPN
12	305	16.7	460	1	PTSB_BACSU
13	299	16.4	699	1	PTGA_BACSU
14	296.5	16.3	456	1	PTSB_SALTY
15	295	16.2	648	1	PTAA_ECOLI
16	294.5	16.2	324	1	PTGA_BACST
17	293.5	16.1	456	1	PTSB_KLEPN
18	273	15.0	459	1	SACK_BACSU
19	270.5	14.9	631	1	YBFS_BACSU
20	256	14.1	189	1	PTGA_BORBU
21	255	14.0	168	1	YPOE_BACSU
22	252.5	13.9	634	1	LACT_STRTR
23	249.5	13.7	154	1	PTGA_MYCCA
24	249.5	13.7	473	1	PTTB_ECOLI
25	246	13.5	168	1	PTBA_ECOLI
26	244	13.4	168	1	PTGA_SALTY
27	238.5	13.1	165	1	PTGA_HAEIN
28	231.5	12.7	161	1	PTGA_BUCAI
29	221	12.1	641	1	RAPF_PEDPE
30	219.5	12.1	627	1	LACT_LACDE
31	207.5	11.4	940	1	PTGA_MYCPN
32	198.5	10.8	908	1	PTGA_MYCPG
33	165.5	9.1	485	1	PTDA_ECOLI

34	161.5	8.9	482	1	PTSB_VIBCH
35	157.5	8.6	639	1	LACY_LEULA
36	151.5	8.3	474	1	YFEV_ECOLI
37	115.5	6.3	538	1	TCMA_STRGA
38	113.5	6.2	1179	1	CIAD_BACTA
39	110	6.0	1951	1	CIN3_HUMAN
40	109	6.0	414	1	FABF_VIBHA
41	107.5	5.9	2249	1	OMPA_RICRI
42	104	5.7	451	1	PTCC_BACST
43	104	5.7	577	1	PTLB_LACCA
44	101.5	5.6	413	1	FABF_VIBCH
45	101.5	5.6	1654	1	OMPB_RICRI

ALIGNMENTS

RESULT	1
PTSA_STRMU	STANDARD; PRT; 664 AA.
AC	P12655;
DT	01-OCT-1989 (Rel. 12, Created)
DT	01-OCT-1989 (Rel. 12, Last sequence update)
DT	15-JUL-1999 (Rel. 38, Last annotation update)
DE	PTS SYSTEM, SUCROSE-SPECIFIC IIABC COMPONENT (ETIABC-SCR) (SUCROSE-DE PERMEASE IIABC COMPONENT) (PHOSPHOTRANSFERASE ENZYME II, ABC COMPONENT) (EC 2.7.1.69) (EII-SCR).
GN	SCRA.
OS	Streptococcus mutans.
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC	Streptococcus.
OX	NCBI_TaxID=1309;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=GS-5;
RC	MEDLINE=89123027; PubMed=2536656;
RA	Sato Y., Poy F., Jacobson G.R., Kuramitsu H.K.;
RT	"Characterization and sequence analysis of the scrA gene encoding enzyme IIScr of the Streptococcus mutans"
RT	phosphoenolpyruvate-dependent sucrose phosphotransferase system.";
RL	J. Bacteriol. 171:263-271(1989).
RL	[2]
RP	SEQUENCE OF 639-664 FROM N.A.
RC	STRAIN=GS-5;
RC	MEDLINE=93329360; PubMed=8336109;
RA	Sato Y., Yamamoto Y., Kizaki H., Kuramitsu H.K.;
RT	"Isolation, characterization and sequence analysis of the scrK gene encoding fructokinase of Streptococcus mutans.";
RL	J. Gen. Microbiol. 139:921-927(1993).
CC	!- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE -TRANSPORT SYSTEM. THE IICD DOMAINS CONTAIN THE SUGAR BINDING SITE AND THE TRANSMEMBRANE CHANNEL; THE IIA DOMAIN CONTAINS THE PRIMARY PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HPR); IIA TRANSFERS ITS PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO THE SUGAR.
CC	!- CATALYTIC ACTIVITY: PROTEIN N-PHOSPHOHISTIDINE + SUGAR -> PROTEIN HISTIDINE + SUGAR PHOSPHATE.
CC	!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC	!- SIMILARITY: CONTAINS A PTS EIIA DOMAIN.
CC	!- SIMILARITY: CONTAINS A PTS EIIIB DOMAIN.
CC	!- SIMILARITY: CONTAINS A PTS EIIIC DOMAIN.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL; M22711; AAA26971.1; -
DR	EMBL; D13175; BAA02466.1; -

```
DR PIR; B32243; B32243.
DR HSSP; P08837; 2P3G.
DR InterPro; IPR001127; PTS_EIIA.
DR InterPro; IPR001996; PTS_EIIB.
DR InterPro; IPR003352; PTS_EIIC.
DR Pfam; PF003358; PTS_EIIA_1; 1.
DR Pfam; PF003367; PTS_EIIB; 1.
DR Pfam; PF02378; PTS_EIIC; 1.
DR ProDom; PD001476; PTS_EIIB; 1.
DR ProDom; PD002243; PTS_EIIA; 1.
DR ProSite; PS00371; PTS_EIIA_1; 1.
DR ProSite; PS01035; PTS_EIIB_Cys; 1.
KW Phosphotransferase system; Sugar transport; Transferase;
KW Phosphorylation; Transmembrane.
FT DOMAIN 1 40
FT DOMAIN ? 2
FT DOMAIN 533 664
FT MOD_RES 25
FT MOD_RES 331 331
FT MOD_RES 585 585
SQ SEQUENCE 664 AA; 69988 MW; 809B63E32281A9A1 CRC64;

Query Match 39.0%; Score 711; DB 1; Length 664;
Best Local Similarity 42.1%; Pred. No. 3.8e-43;
Matches 155; Conservative 73; Mismatches 116; Indels 24; Gaps 8;

QY 1 YDFGGPVGGLLFGVLVSPVITGLHQSFPPIELEL---FNQG---GSFIATASMANIAQ 54
Db 307 YDTTGFGLMGVFGALYSPVVTGLHQSFPPIELEL---FNQG---GSFIATASMANIAQ 366

QY 55 GAACLAFFFLAKSEKLGAGAGSVAVLGTETPAIFGVNLRWLPFFIGTAAIGCAL 114
Db 367 GAATFAIYELTKDKMKGLSSSGVSALLGITETPALGVNLYKFFPFCALIGSASAAI 426

QY 115 IALENIKAVLGAAGFLGVVSDADPMVFLVCAVVTFTAFGAIAIYGLVLRNNGSID 174
Db 427 AGLQVAVVSLGSAFGFLGFSIKASSIPFYVVCLELFAFAFYTYGYG----KTKAVD 481

QY 175 PDATAAPVPAGTTKAEAEAEAFNSDST-----IIOAPLTGPAIALSSVSDAMFASGK 227
Db 482 VFRAEAAREAEAEVQ-EIPEEAASAANKAQVTDVLAAPLAGEAVELTSVNDPVFSEA 540

QY 228 LGSQVAIVPTKGQLVSPVSKIVVAFPSGHAFVTRKAEDGNSVDILMHGFDTVNLNGT 287
Db 541 MGKGIATKSGNTVYAPVDGTQVAFDTGHAYGI--KSDNGA--EILIHIGIDTVSMEGK 596

QY 288 HFNPLKQGVKAGELLCEFDIDAIAKAGYEVTTPIVVSNNYKKGTPVNTY-GLGEIEAG 346
Db 597 GFQKVQADQKIRKGOVLGTGFDSDKIAEAGLDNTTMEFIVTNTADYASVETLASSGTVAVG 656

QY 347 ANLINVAK 354
Db 657 DLSLEVKK 664

RESULT 2
PTSA_PEDPE
ID PTSA_PEDPE STANDARD; PRT; 651 AA.
AC P43470;
DT 01-NOV-1995 (Rel. 32, Created)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE PTS SYSTEM, SUCROSE-SPECIFIC ITABC COMPONENT (ETIABC-SCR) (SUCROSE-
DE PERMEASE ITABC COMPONENT) (PHOSPHOTRANSFERASE ENZYME II, ABC
DE COMPONENT) (EC 2.7.1.69) (EII-SCR).
GN SCRA.
OS Pedicoccus pentosaceus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
OC Pedicoccus.
OX NCBI_TaxID=1255;
RN [1]
RP SEQUENCE FROM N.A.
```

```
RC STRAIN=PEEL.0;
RA Leenbouts K.K.J., Bolhuis A.A., Kok J.J., Venema G.G.;
RL Submitted (XXI-1994) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT
CC SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE
CC -TRANSPORT SYSTEM. THE IID DOMAINS CONTAIN THE SUGAR BINDING-SITE
CC AND THE TRANSMEMBRANE CHANNEL; THE IIA DOMAIN CONTAINS THE PRIMARY
CC PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HPR); IIA TRANSFERS ITS
CC PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO
CC THE SUGAR.
CC -1- CATALYTIC ACTIVITY: PROTEIN N-PHOSPHOHISTIDINE + SUGAR =
CC PROTEIN HISTIDINE + SUGAR PHOSPHATE.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: CONTAINS A PTS EIIA DOMAIN.
CC -1- SIMILARITY: CONTAINS A PTS EIIB DOMAIN.
CC -1- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; Z32771; CAA83668.1; -.
CC HSSP; P20166; IAX3.
CC InterPro; IPR001127; PTS_EIIA.
CC InterPro; IPR001996; PTS_EIIB.
CC InterPro; IPR003352; PTS_EIIC.
CC Pfam; PF003358; PTS_EIIA_1; 1.
CC Pfam; PF003367; PTS_EIIB; 1.
CC Pfam; PF02378; PTS_EIIC; 1.
CC ProDom; PD001476; PTS_EIIB; 1.
CC ProDom; PD002243; PTS_EIIA; 1.
CC ProSite; PS00371; PTS_EIIA_1; 1.
CC ProSite; PS01035; PTS_EIIB_Cys; 1.
KW Phosphotransferase system; Sugar transport; Transferase;
KW Phosphorylation; Transmembrane.
FT DOMAIN 1 40
FT DOMAIN ? 2
FT DOMAIN 510 651
FT MOD_RES 25 25
FT MOD_RES 324 324
FT MOD_RES 562 562
SQ SEQUENCE 651 AA; 68454 MW; C87BA09D550A77F8 CRC64;

Query Match 38.8%; Score 706.5; DB 1; Length 651;
Best Local Similarity 42.0%; Pred. No. 7.6e-43;
Matches 154; Conservative 64; Mismatches 126; Indels 23; Gaps 5;

QY 1 YDFGGPVGGLLFGVLVSPVITGLHQSFPPIELEL-----NQGSFIFATASMANIAQGA 56
Db 300 YNSTGWTGMGIFGLLYSAIVITGLHQTTPAETQTLLANVAKTGTGSGTFPPVSMANIQGA 359

QY 57 ACLAVFFLAKSEKLGAGAGSVAVLGTETPAIFGVNLRWLPFFIGTAAIGGALIA 116
Db 360 ATLAIFFATKSKOKALTSSAGVSALLGITETPAIFGVNLRWLPFFIPFAATASGIAFLG 419

QY 117 LFNIKAVLGAAGFLGVVSDADPMVFLVCAVVTFTAFGAIAIYGLVLRNNGSIDPD 176
Db 420 LFHVLVAMGPASVIGFISTASKSIPAFMLSAVISFVVAIPFIY----AKRTLGDDR 475

QY 177 ATAAPVPAGTTKAEAEAEAFNSDSTIIQAPLTGPAIALSSVSDAMFASGLGVAIVP 236
Db 476 QVKSAPTSTV-----INVNDEIISAPVTGASESLKQVNDQVFSAEIMGKGAIVP 526

QY 237 TKGQLVSPVSKIVVAFPSGHAFVTRKAEDGNSVDILMHGFDTVNLNCTHENPLKKG 296
Db 527 SDSQVVAPADGVITVTDYDSSHAYGIKTTA-----GAELIHGLDITVNLNGEHFTTNNVQK 582
```



QY 357 AVPA 360  
 DB 604 ALLA 607

RESULT 4  
 PTBA\_ERWCH STANDARD; PRT; 631 AA.

AC P26207;  
 DT 01-MAY-1992 (Rel. 22, Created)  
 DT 01-MAY-1992 (Rel. 22, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE PTS SYSTEM, BETA-GLUCOSIDES-SPECIFIC IIABC COMPONENT (EIIABC-BGL)  
 DE (BETA-GLUCOSIDES-PERMEASE IIABC COMPONENT) (PHOSPHOTRANSFERASE  
 DE ENZYME II, ABC COMPONENT) (EC 2.7.1.69) (EII-BGL).  
 GN ARBF.  
 OS Erwinia chrysanthemi.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Pectobacterium.  
 OX NCBI\_TaxID=556;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92121114; PubMed=1732212;  
 RA el Hassouni M., Henrissat B., Chippaux M., Barras F.;  
 RT "Nucleotide sequences of the arb genes, which control beta-glucoside  
 utilization in Erwinia chrysanthemi: comparison with the Escherichia  
 coli bgl operon and evidence for a new beta-glycohydrolase family  
 including enzymes from eubacteria, archaeobacteria, and humans";  
 RL J. Bacteriol. 174:765-777(1992).  
 CC -!- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT  
 SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE  
 -TRANSPORT SYSTEM. THE IICD DOMAINS CONTAIN THE SUGAR BINDING SITE  
 AND THE TRANSMEMBRANE CHANNEL; THE IIA DOMAIN CONTAINS THE PRIMARY  
 PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HPR); IIA TRANSFERS ITS  
 PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO  
 THE SUGAR.  
 CC -!- FUNCTION: ACTS AS BOTH A KINASE AND A PHOSPHATASE ON ARBG (BY  
 SIMILARITY).  
 CC -!- CATALYTIC ACTIVITY: PROTEIN N-PHOSPHOISTIDINE + SUGAR =  
 PROTEIN HISTIDINE + SUGAR PHOSPHATE.  
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE.  
 CC -!- SIMILARITY: CONTAINS A PTS EIIA DOMAIN.  
 CC -!- SIMILARITY: CONTAINS A PTS EIIIB DOMAIN.  
 CC -!- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M81772; AAA24814.1; .  
 DR PIR; B42603; B42603.  
 DR HSP; P20166; IAX3.  
 DR InterPro; IPR001127; PTS\_EIIA.  
 DR InterPro; IPR001996; PTS\_EIIB.  
 DR InterPro; IPR003352; PTS\_EIIC.  
 DR Pfam; PF00358; PTS\_EIIA\_1; 1.  
 DR Pfam; PF00367; PTS\_EIIB; 1.  
 DR Pfam; PF02378; PTS\_EIIC; 1.  
 DR ProDom; PD001476; PTS\_EIIB; 1.  
 DR ProDom; PD002243; PTS\_EIIA; 1.  
 DR ProSITE; PS00371; PTS\_EIIA\_1; 1.  
 DR ProSITE; PS01035; PTS\_EIIB\_Cys; 1.  
 KW Phosphotransferase system; Sugar transport; Transferase;  
 KW Phosphorylation; Transmembrane; Inner membrane.  
 FT DOMAIN 1 41 EIIB DOMAIN.  
 FT DOMAIN ? ? EIIC DOMAIN.  
 FT DOMAIN 501 631 EIIA DOMAIN.  
 FT MOD\_RES 26 26  
 FT MOD\_RES 309 309 PHOSPHORYLATION (BY SIMILARITY).  
 FT MOD\_RES 309 309 PHOSPHORYLATION (BY SIMILARITY).

FT MOD\_RES 553 553 PHOSPHORYLATION (BY SIMILARITY).  
 FT TRANSMEM 50 70 POTENTIAL.  
 FT TRANSMEM 99 119 POTENTIAL.  
 FT TRANSMEM 120 140 POTENTIAL.  
 FT TRANSMEM 146 166 POTENTIAL.  
 FT TRANSMEM 175 195 POTENTIAL.  
 FT TRANSMEM 206 226 POTENTIAL.  
 FT TRANSMEM 248 268 POTENTIAL.  
 FT TRANSMEM 272 292 POTENTIAL.  
 FT TRANSMEM 295 315 POTENTIAL.  
 FT TRANSMEM 328 348 POTENTIAL.  
 FT TRANSMEM 358 378 POTENTIAL.  
 FT TRANSMEM 385 405 POTENTIAL.  
 FT TRANSMEM 407 427 POTENTIAL.  
 FT TRANSMEM 434 454 POTENTIAL.  
 SQ SEQUENCE 631 AA; 66984 MW; 7D0BD27A36BFFDF CRC64;

Query Match 27.0%; Score 492.5; DB 1; Length 631;  
 Best Local Similarity 34.9%; Pred. No. 1e-27;  
 Matches 123; Conservative 62; Mismatches 154; Indels 13; Gaps 7;

QY 1 YFEGGPGVGLLFGLVSPVITGLHQSFPPIELEFNQ--GGSFIFATASMANIAQAACL 59  
 DB 285 YGLNSLLAGAVMGALQVQCVIFGLHWGFVPLMLNNEFVIGHDTLLPLLVPAVLQAGATL 344  
 QY 60 AVFELAKSEKLKLAGASGSAVLGITEPAIFGVNLRWRPFFIGITGAIGALLALFN 119  
 DB 345 GVLLRTQDLKRKGIAAGSAFAIFGITEPAVIGVTPLRRPFFIGGIGGALGAAGVYAH 404  
 QY 120 IKAVALGAAGFLGVWSIDAPDMVFLV--CAVYTFITAFG--AAIAYGLYLVRNRNGSIDPDA 177  
 DB 405 TTMYSFGFFSFTQVPIPTGVDSSVMAAVIGTLLAFAFAALTSNFGVPKD---ETQP 461  
 QY 178 TRAPVPAGTTKAEAPAEFNSDSTIIQAPLGEATALLSVSDAMFASGKLGSGVAIVPT 237  
 DB 462 AAADSPA--VLAETQANAGAVRDETLLF--SPLAGEVLLLEQVADRTFAFGVMGKGIARPT 518  
 QY 238 KGVLPVSGVKIVAFPSGHAFVRKAEDEGNSVDILMHIGEDTVNLNGTHRNPLKKOGD 297  
 DB 519 QGRLYAPVDGTVASLTKTHAIGLASR---GGAEVLIHVIGDITVRLDGRYTFPHVRVGD 574  
 QY 298 EVKAGELLCEDFDIDAKAAGYEVTTPPIVVSNYKTKTGPVNTYGLGEIEAGANL 349  
 DB 575 VVRQGLDLEFDGPALEAAGYDLTTPITVNTSEYDGRGVEPVASGVKVDANAPL 626

RESULT 5  
 PTBA\_ECOLI STANDARD; PRT; 625 AA.

AC P08722;  
 DT 01-AUG-1988 (Rel. 08, Created)  
 DT 01-AUG-1988 (Rel. 08, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE PTS SYSTEM, BETA-GLUCOSIDES-SPECIFIC IIABC COMPONENT (EIIABC-BGL)  
 DE (BETA-GLUCOSIDES-PERMEASE IIABC COMPONENT) (PHOSPHOTRANSFERASE  
 DE ENZYME II, ABC COMPONENT) (EC 2.7.1.69) (EII-BGL).  
 GN BGLF OR BGLC OR BGLS OR B3722.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=88009877; PubMed=3309161;  
 RA Branley H.F., Kornberg H.L.;  
 RT "Nucleotide sequence of bglC, the gene specifying enzymelbgl of the  
 PEP:sugar phosphotransferase system in Escherichia coli K12, and  
 overexpression of the gene product";  
 RL J. Gen. Microbiol. 133:563-573(1987).  
 RN [2]  
 RP SEQUENCE FROM N.A.



```

CC CC -!- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC DR EMBL; X69800; CAA9461.1; -.
CC DR InterPro; IPR001996; PTS_EIIB.
CC DR InterPro; IPR003352; PTS_EIIC.
CC DR Pfam; PF00367; PTS_EIIB; 1.
CC DR Pfam; PF02378; PTS_EIIC; 1.
CC DR ProDom; PD001476; PTS_EIIB; 1.
CC DR PROSITE; PS01035; PTS_EIIB_CYS; 1.
CC KW Phosphotransferase system; Sugar transport; Transferase;
CC KW Transmembrane; Inner membrane; Phosphorylation.
CC FT DOMAIN 1 ? EIIB DOMAIN.
CC FT MOD_RES 26 480
CC FT MOD_RES 325 325 PHOSPHORYLATION (BY SIMILARITY).
CC FT MOD_RES 325 325 PHOSPHORYLATION (BY SIMILARITY).
CC SQ SEQUENCE 480 AA; 51326 MW; AB4E1D9785D84E47 CRC64;

Query Match 24.5%; Score 446.5; DB 1; Length 480;
Best Local Similarity 52.5%; Pred. No. 1.4e-24;
Matches 93; Conservative 30; Mismatches 47; Indels 7; Gaps 3;

QY 1 YDFGGPVGGLFLGVYSPVITGLHQSFPPIELF-----NOGGSPIFATASMANIAQGA 56
Db 301 YFEGGAIGGLIFGLLYAPIVTGMHHSFIAETQLIADSSSTGSGSFIPFIATMSNIAQGA 360

QY 57 ACLAVFFFLAK-SEKLGKLAGAGSVAVLIGTEPAIFGVNLRWPPFIFIGTAAIGGALI 115
Db 361 AALAAFFIIKENKLGKAGSAAAGSALLGITPEPAMFGVNLKLRYPFISGSGISYI 420

QY 116 ALFNKAVAGLGAAGFLGVVSIDAPD--MYMFLVCVVTFEIAFGRAIYGLVLVREN 170
Db 421 AFEKVAIALGAGTAGIPGFIISGQNGNWUHYGIAMIIAFIVAGVTYALSYRKKYRN 477

RESULT 7
IDB PASMU STANDARD; PRT; 474 AA.
AC Q9CJZ2.
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PTS SYSTEM, SUCROSE-SPECIFIC IIBC COMPONENT (EIIB-SCR) (SUCROSE-
DE PERMEASE IIBC COMPONENT) (PHOSPHOTRANSFERASE ENZYME II, BC COMPONENT)
DE (EC 2.7.1.69) (EI-SCR).
GN SCRA OR PTSB OR PM1846.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Pasteurella.
OX NCBI_TaxID=747;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PM70;
RA MEDLINE=21145866; PubMed=11248100;
RX May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
RT "Complete genomic sequence of Pasteurella multocida Pm70.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
CC -!- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT
CC SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE
CC -TRANSPORT SYSTEM. THE IID DOMAINS CONTAIN THE SUGAR BINDING SITE
CC AND THE TRANSMEMBRANE CHANNEL. THE IIA DOMAIN CONTAINS THE PRIMARY
CC PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HPR); IIA TRANSFERS ITS
CC PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO
CC THE SUGAR (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: PROTEIN N-PHOSPHOHISTIDINE + SUGAR =

```

```

CC CC PROTEIN HISTIDINE + SUGAR PHOSPHATE.
CC CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
CC CC (BY SIMILARITY).
CC CC -!- SIMILARITY: CONTAINS A PTS EIIB DOMAIN.
CC CC -!- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC DR EMBL; AE006222; AAK03930.1; -.
CC DR InterPro; IPR001996; PTS_EIIB.
CC DR InterPro; IPR003352; PTS_EIIC.
CC DR Pfam; PF00367; PTS_EIIB; 1.
CC DR Pfam; PF02378; PTS_EIIB; 1.
CC DR PROSITE; PS01035; PTS_EIIB_CYS; FALSE_NEG.
CC KW Phosphotransferase system; Transferase; Sugar transport; Plasmid;
CC KW Transmembrane; Inner membrane; Phosphorylation; Complete proteome.
CC FT DOMAIN 1 ? EIIB DOMAIN.
CC FT DOMAIN 109 474
CC FT TRANSMEM 109 129
CC FT TRANSMEM 158 178
CC FT TRANSMEM 182 202
CC FT TRANSMEM 229 249
CC FT TRANSMEM 284 284
CC FT TRANSMEM 303 323
CC FT TRANSMEM 345 365
CC FT TRANSMEM 376 396
CC FT TRANSMEM 403 423
CC FT TRANSMEM 444 464
CC FT MOD_RES 26 26
CC FT MOD_RES 324 324 PHOSPHORYLATION (BY SIMILARITY).
CC SQ SEQUENCE 474 AA; 49908 MW; D26FA1A059603C01 CRC64;

Query Match 23.0%; Score 418; DB 1; Length 474;
Best Local Similarity 50.9%; Pred. No. 1.4e-22;
Matches 85; Conservative 31; Mismatches 47; Indels 4; Gaps 1;

QY 1 YDFGGPVGGLFLGVYSPVITGLHQSFPPIELFQ-----GGSFIFATASMANIAQGA 56
Db 300 YDTLGFVGGAIFGLTYAPIVTGMHQTFAVETQLLAELAVARTGGTFFIPFIAMSNIAQGA 359

QY 57 ACLAVFFFLAKSEKLGKLAGAGSVAVLIGTEPAIFGVNLRWPPFIFIGTAAIGGALIA 116
Db 360 ACLGAAYVMKDAKVRGIAVPSGISALLGITPEPAMFGVNLRYRYPFISAMIGAGISSAVIA 419

QY 117 LFNKAVAGLGAAGFLGVVSIDAPDVMVFLVCVVTFEIAFGRAIYAG 163
Db 420 LFNKAVIALGAGLPGIPSIRKIPDLSLAMYCVGMLISASIAFTLTVLIG 466

RESULT 8
IDB PASMU STANDARD; PRT; 479 AA.
AC P22825;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PTS SYSTEM, SUCROSE-SPECIFIC IIBC COMPONENT (EIIB-SCR) (SUCROSE-
DE PERMEASE IIBC COMPONENT) (PHOSPHOTRANSFERASE ENZYME II, BC COMPONENT)
DE (EC 2.7.1.69) (EI-SCR).
GN SCRA.
OS Vibrio alginolyticus.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=663;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91071601; PubMed=2174811;

```

Blatch G.L., Scholle R.R., Woods D.R.;  
 "Nucleotide sequence and analysis of the *Vibrio alginolyticus* sucrose  
 uptake-encoding region.";  
 RL Gene 95:17-23(1990).  
 [2]  
 RN  
 RP SEQUENCE OF 1-12 FROM N.A.  
 RX MEDLINE=91285433; PubMed=2060795;  
 RA Blatch G.L., Woods D.R.;  
 RT "Nucleotide sequence and analysis of the *Vibrio alginolyticus* scr  
 repressor-encoding gene (scrR).";  
 RL Gene 101:45-50(1991).  
 CC  
 CC -!- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT  
 SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE  
 -TRANSPORT SYSTEM. THE IICD DOMAINS CONTAIN THE SUGAR BINDING SITE  
 AND THE TRANSMEMBRANE CHANNEL; THE IIA DOMAIN CONTAINS THE PRIMARY  
 PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HPR); IIA TRANSFERS ITS  
 PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO  
 THE SUGAR.  
 CC  
 CC -!- CATALYTIC ACTIVITY: PROTEIN N-PHOSPHOHISTIDINE + SUGAR =  
 PROTEIN HISTIDINE + SUGAR PHOSPHATE.  
 CC  
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE. INNER MEMBRANE.  
 CC  
 CC -!- SIMILARITY: CONTAINS A PTS EIIB DOMAIN.  
 CC  
 CC -!- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 or send an email to license@isb-sib.ch).  
 CC  
 CC EMBL; M76768; AAA27555.1; -  
 DR EMBL; M35009; AAA27557.2; -  
 DR PIR; JQ0781; JQ0781.  
 DR HSSP; P05053; IIBA.  
 DR InterPro; IPR001996; PTS\_EIIB.  
 DR InterPro; IPR003352; PTS\_EIIC.  
 DR Pfam; PF00367; PTS\_EIIB; 1.  
 DR Pfam; PF02378; PTS\_EIIC; 1.  
 DR ProDom; PD001476; PTS\_EIIB; 1.  
 DR PROSITE; PS01035; PTS\_EIIB-CYS; 1.  
 KW Phosphotransferase system; Sugar transport; Transferase;  
 KW Transmembrane; Inner membrane; Phosphorylation.  
 FT DOMAIN 1 ? EIIB DOMAIN.  
 FT DOMAIN ? 479 EIIC DOMAIN.  
 FT MOD\_RES 26 26 PHOSPHORYLATION (BY SIMILARITY).  
 FT MOD\_RES 324 324 PHOSPHORYLATION (BY SIMILARITY).  
 SQ SEQUENCE 479 AA; 49890 MW; FBF906B5170E3EB7 CRC64;  
 Query Match 22.6%; Score 412; DB 1; Length 479;  
 Best Local Similarity 49.7%; Pred. No. 3.8e-22;  
 Matches 84; Conservative 32; Mismatches 47; Indels 6; Gaps 3;  
 Qy 1 YDFGPGVGLLFLGLVYSPVITLGLHSPPIELFLF----NOGGSFIFATASMANIAQGA 56  
 Db 300 YNTAGVGVGFLIYAPVIGMHHSFATIAETQLADIATTGGTFIPPIAAMSNVSGA 359  
 Qy 57 ACLAVFFLAKSEKLGAGASGVAVLGPTEPAIFGVNLRWLPFFIGITGAIGGALIA 116  
 Db 360 AALAVGVMSKDKMKGIAPSGVTGLLGITPEPAMGVNLRWLPFIAPVAAALSSAFIT 419  
 Qy 117 LFNKAVAGAGFGLGVVSDIDPDMWMLVCA-VVTFITAFGAAYGL 164  
 Db 420 MENVKAQALGAAGLGLIISI-TPDKIGVYIAGWIAFLTAFLVTLVGI 467  
 RESULT 9  
 ID PTGA\_CORGL STANDARD; PRT; 674 AA.  
 AC Q45298;  
 DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE PTS SYSTEM, GLUCOSE-SPECIFIC IIB COMPONENT (EIIBC-GLC) (GLUCOSE-  
 DE PERMEASE IIB COMPONENT) (PHOSPHOTRANSFERASE ENZYME II, ABC  
 DE COMPONENT) (EC 2.7.1.69) (EII-GLC/EIIB-GLC).  
 OS PTSG.  
 OS Corynebacterium glutamicum (Brevibacterium flavum).  
 OG Plasmid pBSBG2.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Corynebacterineae; Corynebacteriaceae;  
 OC Corynebacterium.  
 OX NCBI\_TaxID=17118;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 13869;  
 RA Yoon K.-H.;  
 RT "Cloning and nucleotide sequence of enzyme II of Brevibacterium  
 lactofermentum phosphotransferase system.";  
 RL Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT  
 SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE  
 -TRANSPORT SYSTEM. THE IICD DOMAINS CONTAIN THE SUGAR BINDING SITE  
 AND THE TRANSMEMBRANE CHANNEL; THE IIA DOMAIN CONTAINS THE PRIMARY  
 PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HPR); IIA TRANSFERS ITS  
 PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO  
 THE SUGAR.  
 CC  
 CC -!- CATALYTIC ACTIVITY: PROTEIN N-PHOSPHOHISTIDINE + SUGAR =  
 PROTEIN HISTIDINE + SUGAR PHOSPHATE.  
 CC  
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
 CC  
 CC -!- SIMILARITY: CONTAINS A PTS EIIB DOMAIN.  
 CC  
 CC -!- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.  
 CC  
 CC -!- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 or send an email to license@isb-sib.ch).  
 CC  
 CC EMBL; L18875; AAA22992.1; -  
 DR HSSP; P08837; 2F3G.  
 DR InterPro; IPR001127; PTS\_EIIB.  
 DR InterPro; IPR001996; PTS\_EIIB.  
 DR InterPro; IPR003352; PTS\_EIIC.  
 DR Pfam; PF00358; PTS\_EIIB; 1.  
 DR Pfam; PF00367; PTS\_EIIB; 1.  
 DR Pfam; PF02378; PTS\_EIIC; 1.  
 DR ProDom; PD001476; PTS\_EIIB; 1.  
 DR PROSITE; PS00371; PTS\_EIIB; 1.  
 DR PROSITE; PS01035; PTS\_EIIB-CYS; 1.  
 KW Phosphotransferase system; Sugar transport; Transferase;  
 KW Phosphorylation; Transmembrane; Plasmid.  
 FT DOMAIN 1 43 EIIB DOMAIN.  
 FT DOMAIN ? ? EIIC DOMAIN.  
 FT DOMAIN 542 674 EIIB DOMAIN.  
 FT TRANSMEM 126 146 POTENTIAL.  
 FT TRANSMEM 162 182 POTENTIAL.  
 FT TRANSMEM 193 213 POTENTIAL.  
 FT TRANSMEM 225 245 POTENTIAL.  
 FT TRANSMEM 260 280 POTENTIAL.  
 FT TRANSMEM 303 323 POTENTIAL.  
 FT TRANSMEM 344 364 POTENTIAL.  
 FT TRANSMEM 377 397 POTENTIAL.  
 FT TRANSMEM 409 429 POTENTIAL.  
 FT TRANSMEM 442 462 POTENTIAL.  
 FT TRANSMEM 493 513 POTENTIAL.  
 FT MOD\_RES 28 28 PHOSPHORYLATION (BY SIMILARITY).  
 FT MOD\_RES 594 594 PHOSPHORYLATION (BY SIMILARITY).  
 SQ SEQUENCE 674 AA; 71626 MW; 67A75AAF76E42FA2 CRC64;

Query Match 21.2%; Score 386.5; DB 1; Length 674;  
Best Local Similarity 31.0%; Pred. No. 3.6e-20;  
Matches 116; Conservative 61; Mismatches 156; Indels 41; Gaps 9;

QY 10 LFLGLVSPVITGLHQSFPPPIELEFNQ--GGSFIFATASMANIAOGAACLAFFVLAKSE 68  
DB 307 IVIPLLYPLVPLGLHWPLNAINQNLNTLGYDFIQGPMGANNFACFGLVGTGFLIALKE 366  
QY 69 KLKGLAGAS--GVSAYL--GITEPATFGVNLRLWRPFETGTAATGGAIALIENIKAVA 124  
DB 367 KNAMRQVSLGGLAGLGGISEPSLYGLVLLKFKTYFRLPLGCLGVGVIMGIFDIKAYA 426  
QY 125 LGAAGLGVVSDAPDMWFLVCAVVTETFAAGAAATAGLYLVRRNGSID----- 174  
DB 427 F---VFTSLTLLTPAMPDPLWGLTYGVIAAFA--FTSMLLVLFFDYRSDAERDEAKAQAAMAE 481  
QY 175 -----PDATAAPV-----PAGTTKAAEAFAEFSNDSTIIQAPLITGSAIALSSV 218  
DB 482 QTNNTPAAPAAPVAPAAAGAAAGGAAGAT-AVATKPRLAAGQLVEITSPLEGHAVPLSEV 540  
QY 219 SDAMFASGLKSGVAIVPTKQLVSPVSGKIVVAFPSGHAFVTRKAEDGNSVDILMHIG 278  
DB 541 PDPIFAAGLPGIAEPTGTNTVAPADATVILVQKSGHVALRLRLE----SGVELLIHIG 596  
QY 279 FDTVNLNGTHFNPLKKGDEVKAGELLCEDIDAIRKAAAGVEVTTPIVWSNYKKTGPVNTY 338  
DB 597 LDTVQLGGEGFVHVVERKQVRAKADPLITFDPEFIRSKNPLITIPVVSNAKFGELVGI 656  
QY 339 GLGEIEAGANLLNV 352  
DB 657 ERAQAADATTVIKV 670

RESULT 10  
ID PTTB\_BACSU STANDARD; PRT; 470 AA.  
AC P39794; 034771;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE PROBABLE PTS SYSTEM, TREHALOSE-SPECIFIC IIBC COMPONENT (EIIBC-TRE)  
DE (TREHALOSE)PEREASE IIBC COMPONENT) (PHOSPHOTRANSFERASE ENZYME II, BC  
DE COMPONENT) (EC 2.7.1.69) (EII-TRE).  
GN TREP OR TREB.  
OS Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
OC Bacillus/Staphylococcus group; Bacillus.  
OX NCBI\_TaxID=1423;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168 / MARBURG;  
RX MEDLINE=97074649; PubMed=8917076;  
RA Schoeck F., Dahl M.K.;  
RT "Analysis of DNA flanking the treA gene of Bacillus subtilis reveals  
RT genes encoding a putative specific enzyme IITRE and a potential  
RT regulator of the trehalose operon.";  
RL Gene 175:59-63(1996).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=AC327;  
RX MEDLINE=97124190; PubMed=8969503;  
RA Yamamoto H., Uchiyama S., Sekiguchi J.;  
RT "Cloning and sequencing of a 40.6 kb segment in the 73 degrees-76  
RT degrees region of the Bacillus subtilis chromosome containing genes  
RT for trehalose metabolism and acetoin utilization.";  
RL Microbiology 142:3057-3065(1996).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=AC327;  
RX MEDLINE=97417488; PubMed=9272861;  
RA Yamamoto H., Uchiyama S., Nugroho F.A., Sekiguchi J.;  
RT "Cloning and sequencing of a 35.7 kb in the 70 degree-73 degree region

of the Bacillus subtilis genome reveal genes for a new two-component  
system, three spore germination proteins, an iron uptake system and a  
general stress response protein.";  
Gene 194:191-199(1997).  
[4]  
RP SEQUENCE OF 324-470 FROM N.A.  
RC STRAIN=168;  
RX MEDLINE=95379486; PubMed=7651129;  
RA Helfert C., Gotsche S., Dahl M.K.;  
RT "Cleavage of trehalose-phosphate in Bacillus subtilis is catalysed by  
RT a phospho-alpha-(1-1)-glucosidase encoded by the treA gene.";  
RL Mol. Microbiol. 16:111-120(1995).  
CC -!- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT  
CC SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE  
CC -TRANSPORT SYSTEM. THE IICD DOMAINS CONTAIN THE SUGAR BINDING SITE  
CC AND THE TRANSMEMBRANE CHANNEL; THE IIA DOMAIN CONTAINS THE PRIMARY  
CC PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HPR); IIA TRANSFERS ITS  
CC PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO  
CC THE SUGAR.  
CC -!- CATALYTIC ACTIVITY: PROTEIN N-PHOSPHO-HISTIDINE + SUGAR =  
CC PROTEIN HISTIDINE + SUGAR PHOSPHATE.  
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE.  
CC -!- SIMILARITY: CONTAINS A PTS EIIIB DOMAIN.  
CC -!- SIMILARITY: CONTAINS A PTS EIIIC DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
DR EMBL; Z54245; CAA91014.1; -;  
DR EMBL; D83967; BAA23409.1; -;  
DR EMBL; D86417; BAA22289.1; -;  
DR EMBL; X80203; CAA56494.1; -;  
DR EMBL; Z99108; CAB12609.1; -;  
DR HSSP; P05053; 1IBA.  
DR Subtilist; BG11009; trep.  
DR InterPro; IPR001996; PTS\_EIIB.  
DR InterPro; IPR003352; PTS\_EIIC.  
DR Pfam; PF00367; PTS\_EIIB; 1.  
DR Pfam; PF02378; PTS\_EIIC; 1.  
DR ProDom; PD001476; PTS\_EIIB; 1.  
DR PROSITE; PS01035; PTS\_EIIB\_CYS; 1.  
KW Phosphotransferase system; Sugar transport; Transferase;  
KW Transmembrane; Inner membrane; Phosphorylation; Complete proteome.  
FT DOMAIN 1 ?  
FT DOMAIN ? 470  
FT MOD\_RES 26 26 PHOSPHORYLATION (BY SIMILARITY).  
FT TRANSMEM 110 130 POTENTIAL.  
FT TRANSMEM 160 180 POTENTIAL.  
FT TRANSMEM 183 203 POTENTIAL.  
FT TRANSMEM 234 254 POTENTIAL.  
FT TRANSMEM 263 283 POTENTIAL.  
FT TRANSMEM 301 321 POTENTIAL.  
FT TRANSMEM 326 346 POTENTIAL.  
FT TRANSMEM 347 367 POTENTIAL.  
FT TRANSMEM 375 395 POTENTIAL.  
FT TRANSMEM 403 423 POTENTIAL.  
FT TRANSMEM 443 463 POTENTIAL.  
FT CONFLICT 140 140 F -> S (IN REF. 1).  
FT CONFLICT 363 363 M -> L (IN REF. 1 AND 4).  
FT CONFLICT 465 465 A -> G (IN REF. 1 AND 4).  
SQ SEQUENCE 470 AA; 49999 MW; 7A741850A2697D53 CRC64;

Query Match 18.1%; Score 330; DB 1; Length 470;  
Best Local Similarity 41.6%; Pred. No. 2.4e-16;  
Matches 67; Conservative 34; Mismatches 58; Indels 2; Gaps 1;

QY 7 VGLLGLVSPVITGLHQSFPPPIELEFNQ--GGSFIFATASMANIAOGAACLAFFVL 64



Db 307 LGGLYGGFYSAVITGMHTTFLAVDLQIGSKIGTFLWPMIALSNIAQSGSAALAMFI 366  
 QY 65 AKSEKLGAGSAGSVAVLIGTEPAIFGNVLRWPPFFIGITGAIGGALIALFNKAVA 124  
 Db 367 VKDERQKGLSLTSGISAYLIGTEPAIFGNVLRWPPFFIAYSSGLAGNYISSQGLASS 426  
 QY 125 LCAAGCFLGVVSDPDMVNFCAVVTFFIAFGAIAAYGLY 165  
 Db 427 VGVGVPGLFISMSQYGAFAIGMAIVLIVPAGIYAYARF 467

## RESULT 11

PTAA\_KLEPN  
 ID PTAA\_KLEPN STANDARD; PRT; 651 AA.  
 AC P45604;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE PTS SYSTEM, N-ACETYLGLUCOSAMINE-SPECIFIC IIBAC COMPONENT (EIIABC-NAG)  
 DE (N-ACETYLGLUCOSAMINE-PERMEASE IIBAC COMPONENT) (PHOSPHOTRANSFERASE;  
 DE ENZYME II, ABC COMPONENT) (EC 2.7.1.69) (EII-NAG).  
 GN NAG.

OS Klebsiella pneumoniae.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Klebsiella.

OX NCBI\_TaxID-573;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RC STRAIN-1033-5p14 / KAY2026;  
 RX MEDLINE-92079906; PubMed-1745234;

RA Vogler A.P., Lengeler J.W.;

RT "Comparison of the sequences of the nagE operons from Klebsiella pneumoniae and Escherichia coli K12: enhanced variability of the enzyme IIN-acetylglucosamine in regions connecting functional domains";  
 RT Mol. Gen. Genet. 230:270-276(1991).

CC -!- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE -TRANSPORT SYSTEM. THE IID DOMAINS CONTAIN THE SUGAR BINDING SITE AND THE TRANSMEMBRANE CHANNEL; THE IIA DOMAIN CONTAINS THE PRIMARY PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HPR); IIA TRANSFERS ITS PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO THE SUGAR.

CC -!- CATALYTIC ACTIVITY: PROTEIN N-PHOSPHOTIDINE + SUGAR =

CC PROTEIN HISTIDINE + SUGAR PHOSPHATE.

CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE.

CC -!- SIMILARITY: CONTAINS A PTS EIIA DOMAIN.

CC -!- SIMILARITY: CONTAINS A PTS EIIA DOMAIN.

CC -!- SIMILARITY: CONTAINS A PTS EIIIC DOMAIN.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----

CC EMBL; X63289; CAA44923.1; -

CC HSSP; P08837; 2F3G.

CC InterPro; IPR001127; PTS\_EIIA.

CC InterPro; IPR001996; PTS\_EIIB.

CC InterPro; IPR003352; PTS\_EIIC.

CC Pfam; PF00358; PTS\_EIIA.1; 1.

CC Pfam; PF00367; PTS\_EIIB; 1.

CC Pfam; PF02378; PTS\_EIIC; 1.

CC ProDom; PD001476; PTS\_EIIB; 1.

CC ProDom; PD002243; PTS\_EIIA; 1.

CC PROSITE; PS00371; PTS\_EIIA.1; 1.

CC PROSITE; PS01035; PTS\_EIIB.CYS; 1.

KW Phosphotransferase system; Sugar transport; Transferase;

KW Phosphorylation; Transmembrane; Inner membrane.

FT DOMAIN 1 ?  
 FT DOMAIN 393 427 EIIIC DOMAIN.  
 FT DOMAIN 519 651 EIIIB DOMAIN.  
 FT MOD\_RES 190 190 EIIIA DOMAIN.  
 FT MOD\_RES 412 412 PHOSPHORYLATION (BY SIMILARITY).  
 FT MOD\_RES 571 571 PHOSPHORYLATION (BY SIMILARITY).  
 FT TRANSMEM 16 36 POTENTIAL.  
 FT TRANSMEM 40 60 POTENTIAL.  
 FT TRANSMEM 70 90 POTENTIAL.  
 FT TRANSMEM 92 112 POTENTIAL.  
 FT TRANSMEM 132 152 POTENTIAL.  
 FT TRANSMEM 165 185 POTENTIAL.  
 FT TRANSMEM 192 212 POTENTIAL.  
 FT TRANSMEM 232 252 POTENTIAL.  
 FT TRANSMEM 264 284 POTENTIAL.  
 FT TRANSMEM 285 305 POTENTIAL.  
 FT TRANSMEM 308 328 POTENTIAL.  
 FT TRANSMEM 339 359 POTENTIAL.  
 SQ SEQUENCE 651 AA; 68179 MW; 3D8D8ADF4BD48ED CRC64;

Query Match 17.2%; Score 313; DB 1; Length 651;

Best Local Similarity 23.1%; Pred. No. 5.5e-15;

Matches 113; Conservative 67; Mismatches 144; Indels 166; Gaps 15;

QY 5 GPGVGLLGLVSPVITGLHOSFPPI-----EELFNOG----GSF 42  
 Db 170 GALSGSIFGNLLIPTGLHQVNTIAWFQIGFTNAACTVFHGDINRFYAGDTAGMF 229  
 QY 43 IFATASMANIAQGAACLAFLAKSEK---LKLAGASGVSAVL-GITEPAIF----- 91  
 Db 230 MSGFFPINMFLGPGAALAMYLAAKARRPMVGMGLLSVAITFLTGTETLEFLFLAP 289  
 QY 92 -----GVNLRWRPFFIGITGAAGGAL--IALFNKAVLGAAGFLGVWSIDAP 139  
 Db 290 LLYLLHAVLTGISLTATLGIHAGFSFAGADYVLMYSLPAA-----SK 335  
 QY 140 DMVMEVLCVAVTFEII---AFGAIAIY-----GL----- 164  
 Db 336 NVMLLMVGMVFFVFFVLLFSVIRMFNLKTPGREDKAADVVTEANSTEGLTQIATS 395  
 QY 165 YLVRRNG-----SID-----174  
 Db 396 YIAAVGGTDNLKAIDACITRLRLTVGDSAKVNDAAACKRLGASGVVKNKOTIQVIVGAKA 455  
 QY 175 -----PDATAAPVPAG---TTKAAEAAPAEFSNDSTI--IOAPLTGEIAIALS 216  
 Db 456 ESTGDEMKVWTRGPVAAAAAAGNVAATAAPAAKQPAVANAKTVESLVSPTIGDVVALE 515  
 QY 217 SYSDAMFASGKLGSGVAIVPTKGQVSPVSGKIVVAFPSGHAFVTRKAEKDSNVLDLMH 276  
 Db 516 QVPDEAFASKAVGDDGIAVKPTDNIVVAPAGTVVKIFNTNHAFCLLET-----NNGAEIVVH 571  
 QY 277 IGFDTVNLNGTHFNPLKKGQDEVKAGELCEFDIDAIIKAAGVEVTTPIVVSNNYKKTGPVN 336  
 Db 572 MGDITVALSGKFKRLVEGSDTVKAGEPILEMDLDFLNANARSMISPVVCSNDDISALV 631  
 QY 337 TYGLGIEIAG 346  
 Db 632 ILASGKVAVG 641

## RESULT 12

PTSB\_BACSU  
 ID PTSB\_BACSU STANDARD; PRT; 460 AA.  
 AC P05306;  
 DT 01-NOV-1988 (Rel. 09, Created)  
 DT 01-NOV-1988 (Rel. 09, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE PTS SYSTEM, SUCROSE-SPECIFIC IIBC COMPONENT (EIIIC-SCR) (SUCROSE-  
 DE PERMEASE IIBC COMPONENT) (PHOSPHOTRANSFERASE ENZYME II, BC COMPONENT)  
 DE (EC 2.7.1.69) (EII-SCR).  
 GN SACP OR IPA-49D.

Bacillus subtilis.  
Bacteria; Firmicutes; Bacillus/Clostridium group;  
Bacillus/Staphylococcus group; Bacillus.

OS  
OC  
OX  
RN

SEQUENCE FROM N.A.

MEDLINE=88097369; PubMed=3122206;

Fouet A., Arnaud M., Klier A., Rapoport G.;

"Bacillus subtilis sucrose-specific enzyme II of the

phosphotransferase system: expression in *Escherichia coli* and

homology to enzymes II from enteric bacteria.";

Proc. Natl. Acad. Sci. U.S.A. 84:8773-8777(1987).

[1]

NCBI\_TaxID=1423;

SEQUENCE FROM N.A.

MEDLINE=95020537; PubMed=7934828;

Glaser P., Kunst F., Arnaud M., Coudart M.P., Gonzales W.,

Hullo M.F., Ionescu M., Lubochinsky B., Marcelino L., Mosser I.,

Presecan E., Santana M., Schneider E., Schweizer J., Vertes A.,

Rapoport G., Danchin A.;

"Bacillus subtilis genome project: cloning and sequencing of the 97

kb region from 325 degrees to 333 degrees.";

Mol. Microbiol. 10:371-384(1993).

[2]

FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT

SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE

-TRANSPORT SYSTEM. THE IICD DOMAINS CONTAIN THE SUGAR BINDING SITE

AND THE TRANSMEMBRANE CHANNEL; THE IIA DOMAIN CONTAINS THE PRIMARY

PHOSPHORYLATION SITE (THE DONOR IS PHOSHO-HPR); IIA TRANSFERS ITS

PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO

THE SUGAR.

CC

FUNCTION: ACTS AS BOTH A KINASE AND A PHOSPHATASE ON SACT (BY

SIMILARITY).

CC

CATALYTIC ACTIVITY: PROTEIN N-PHOSPHOHISTIDINE + SUGAR =

PROTEIN HISTIDINE + SUGAR PHOSPHATE.

CC

SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.

CC

SIMILARITY: CONTAINS A PTS EIIB DOMAIN.

CC

SIMILARITY: CONTAINS A PTS EIIC DOMAIN.

CC

SIMILARITY: 57% IDENTITY WITH THE NEGATIVE REGULATORY PROTEIN OF

SACY.

CC

-----

This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC

EMBL; J03006; AAA22727.1; -

DR

EMBL; X73124; CAA51605.1; -

DR

EMBL; Z99123; CAB15831.1; -

DR

PIR; A39938; A39938.

DR

PIR; S39704; S39704.

DR

HSSP; P05053; IIBA.

DR

Subtilast; BG10595; sacP.

DR

InterPro; IPR001996; PTS\_EIIB.

DR

InterPro; IPR003352; PTS\_EIIC.

DR

Pfam; PF00367; PTS\_EIIB; 1.

DR

Pfam; PF02378; PTS\_EIIC; 1.

DR

ProDom; PD001476; PTS\_EIIB; 1.

DR

PROSITE; PS01035; PTS\_EIIB\_CYS; 1.

OS  
OC  
OX  
RN

SEQUENCE FROM N.A.

MEDLINE=88097369; PubMed=3122206;

Fouet A., Arnaud M., Klier A., Rapoport G.;

"Bacillus subtilis sucrose-specific enzyme II of the

phosphotransferase system: expression in *Escherichia coli* and

homology to enzymes II from enteric bacteria.";

Proc. Natl. Acad. Sci. U.S.A. 84:8773-8777(1987).

[1]

NCBI\_TaxID=1423;

SEQUENCE FROM N.A.

MEDLINE=95020537; PubMed=7934828;

Glaser P., Kunst F., Arnaud M., Coudart M.P., Gonzales W.,

Hullo M.F., Ionescu M., Lubochinsky B., Marcelino L., Mosser I.,

Presecan E., Santana M., Schneider E., Schweizer J., Vertes A.,

Rapoport G., Danchin A.;

"Bacillus subtilis genome project: cloning and sequencing of the 97

kb region from 325 degrees to 333 degrees.";

Mol. Microbiol. 10:371-384(1993).

[2]

FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT

SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE

-TRANSPORT SYSTEM. THE IICD DOMAINS CONTAIN THE SUGAR BINDING SITE

AND THE TRANSMEMBRANE CHANNEL; THE IIA DOMAIN CONTAINS THE PRIMARY

PHOSPHORYLATION SITE (THE DONOR IS PHOSHO-HPR); IIA TRANSFERS ITS

PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO

THE SUGAR.

CC

FUNCTION: ACTS AS BOTH A KINASE AND A PHOSPHATASE ON SACT (BY

SIMILARITY).

CC

CATALYTIC ACTIVITY: PROTEIN N-PHOSPHOHISTIDINE + SUGAR =

PROTEIN HISTIDINE + SUGAR PHOSPHATE.

CC

SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.

CC

SIMILARITY: CONTAINS A PTS EIIB DOMAIN.

CC

SIMILARITY: CONTAINS A PTS EIIC DOMAIN.

CC

SIMILARITY: 57% IDENTITY WITH THE NEGATIVE REGULATORY PROTEIN OF

SACY.

CC

-----

This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC

EMBL; J03006; AAA22727.1; -

DR

EMBL; X73124; CAA51605.1; -

DR

EMBL; Z99123; CAB15831.1; -

DR

PIR; A39938; A39938.

DR

PIR; S39704; S39704.

DR

HSSP; P05053; IIBA.

DR

Subtilast; BG10595; sacP.

DR

InterPro; IPR001996; PTS\_EIIB.

DR

InterPro; IPR003352; PTS\_EIIC.

DR

Pfam; PF00367; PTS\_EIIB; 1.

DR

Pfam; PF02378; PTS\_EIIC; 1.

DR

ProDom; PD001476; PTS\_EIIB; 1.

DR

PROSITE; PS01035; PTS\_EIIB\_CYS; 1.

OS  
OC  
OX  
RN

SEQUENCE FROM N.A.

MEDLINE=88097369; PubMed=3122206;

Fouet A., Arnaud M., Klier A., Rapoport G.;

"Bacillus subtilis sucrose-specific enzyme II of the

phosphotransferase system: expression in *Escherichia coli* and

homology to enzymes II from enteric bacteria.";

Proc. Natl. Acad. Sci. U.S.A. 84:8773-8777(1987).

[1]

NCBI\_TaxID=1423;

SEQUENCE FROM N.A.

MEDLINE=95020537; PubMed=7934828;

Glaser P., Kunst F., Arnaud M., Coudart M.P., Gonzales W.,

Hullo M.F., Ionescu M., Lubochinsky B., Marcelino L., Mosser I.,

Presecan E., Santana M., Schneider E., Schweizer J., Vertes A.,

Rapoport G., Danchin A.;

"Bacillus subtilis genome project: cloning and sequencing of the 97

kb region from 325 degrees to 333 degrees.";

Mol. Microbiol. 10:371-384(1993).

[2]

FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT

SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE

-TRANSPORT SYSTEM. THE IICD DOMAINS CONTAIN THE SUGAR BINDING SITE

AND THE TRANSMEMBRANE CHANNEL; THE IIA DOMAIN CONTAINS THE PRIMARY

PHOSPHORYLATION SITE (THE DONOR IS PHOSHO-HPR); IIA TRANSFERS ITS

PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO

THE SUGAR.

CC

FUNCTION: ACTS AS BOTH A KINASE AND A PHOSPHATASE ON SACT (BY

SIMILARITY).

CC

CATALYTIC ACTIVITY: PROTEIN N-PHOSPHOHISTIDINE + SUGAR =

PROTEIN HISTIDINE + SUGAR PHOSPHATE.

CC

SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.

CC

SIMILARITY: CONTAINS A PTS EIIB DOMAIN.

CC

SIMILARITY: CONTAINS A PTS EIIC DOMAIN.

CC

SIMILARITY: 57% IDENTITY WITH THE NEGATIVE REGULATORY PROTEIN OF

SACY.

CC

-----

This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC

EMBL; J03006; AAA22727.1; -

DR

EMBL; X73124; CAA51605.1; -

DR

EMBL; Z99123; CAB15831.1; -

DR

PIR; A39938; A39938.

DR

PIR; S39704; S39704.

DR

HSSP; P05053; IIBA.

DR

Subtilast; BG10595; sacP.

DR

InterPro; IPR001996; PTS\_EIIB.

DR

InterPro; IPR003352; PTS\_EIIC.

DR

Pfam; PF00367; PTS\_EIIB; 1.

DR

Pfam; PF02378; PTS\_EIIC; 1.

DR

ProDom; PD001476; PTS\_EIIB; 1.

DR

PROSITE; PS01035; PTS\_EIIB\_CYS; 1.

OS  
OC  
OX  
RN

SEQUENCE FROM N.A.

MEDLINE=88097369; PubMed=3122206;

Fouet A., Arnaud M., Klier A., Rapoport G.;

"Bacillus subtilis sucrose-specific enzyme II of the

phosphotransferase system: expression in *Escherichia coli* and

homology to enzymes II from enteric bacteria.";

Proc. Natl. Acad. Sci. U.S.A. 84:8773-8777(1987).

[1]

NCBI\_TaxID=1423;

SEQUENCE FROM N.A.

MEDLINE=95020537; PubMed=7934828;

Glaser P., Kunst F., Arnaud M., Coudart M.P., Gonzales W.,

Hullo M.F., Ionescu M., Lubochinsky B., Marcelino L., Mosser I.,

Presecan E., Santana M., Schneider E., Schweizer J., Vertes A.,

Rapoport G., Danchin A.;

"Bacillus subtilis genome project: cloning and sequencing of the 97

kb region from 325 degrees to 333 degrees.";

Mol. Microbiol. 10:371-384(1993).

[2]

FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT

SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE

-TRANSPORT SYSTEM. THE IICD DOMAINS CONTAIN THE SUGAR BINDING SITE

AND THE TRANSMEMBRANE CHANNEL; THE IIA DOMAIN CONTAINS THE PRIMARY

PHOSPHORYLATION SITE (THE DONOR IS PHOSHO-HPR); IIA TRANSFERS ITS

PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO

THE SUGAR.

CC

FUNCTION: ACTS AS BOTH A KINASE AND A PHOSPHATASE ON SACT (BY

SIMILARITY).

CC

CATALYTIC ACTIVITY: PROTEIN N-PHOSPHOHISTIDINE + SUGAR =

PROTEIN HISTIDINE + SUGAR PHOSPHATE.

CC

SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.

CC

SIMILARITY: CONTAINS A PTS EIIB DOMAIN.

CC

SIMILARITY: CONTAINS A PTS EIIC DOMAIN.

RA MEDLINE=91299743; PubMed=1906345;  
RA Fairbrother W.J., Cavanagh J., Dyson H.J., Plamer A.G. III,  
RA Sutrina S.L., Reizer J., Salier M.H. Jr., Wright P.E.;  
RT "Polypeptide backbone resonance assignments and secondary structure  
of Bacillus subtilis iifig determined by two-dimensional and  
three-dimensional heteronuclear NMR spectroscopy";  
RL Biochemistry 30:6896-6907(1991).  
RN [6]  
RP STRUCTURE BY NMR OF EIAA DOMAIN.  
RX MEDLINE=98254029; PubMed=9593197;  
RA Chen Y., Case D.A., Reizer J., Salier M.H. Jr., Wright P.E.;  
RT "High-resolution solution structure of Bacillus subtilis IIAglc.";  
RL Proteins 31:258-270(1998).  
CC -1- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT  
CC SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE  
CC -TRANSPORT SYSTEM. THE IIDC DOMAINS CONTAIN THE SUGAR BINDING SITE  
CC AND THE TRANSMEMBRANE CHANNEL; THE IIA DOMAIN CONTAINS THE PRIMARY  
CC PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HPR); IIA TRANSFERS ITS  
CC PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO  
CC THE SUGAR.  
CC -1- CATALYTIC ACTIVITY: PROTEIN N-PHOSPHOHISTIDINE + SUGAR -  
CC PROTEIN HISTIDINE + SUGAR PHOSPHATE.  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
CC -1- SIMILARITY: CONTAINS A PTS EIIA DOMAIN.  
CC -1- SIMILARITY: CONTAINS A PTS EIIB DOMAIN.  
CC -1- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; Z11744; CA477803.1; -  
DR EMBL; X12832; CA431315.1; -  
DR EMBL; M60344; AA22498.1; -  
DR EMBL; Z99111; CAB13262.1; -  
DR PIR; S22752; WQBSGS.  
DR PDB; 1GPR; 31-OCT-93.  
DR PDB; 1AX3; 17-JUN-98.  
DR Subtilist; BG10198; ptsg.  
DR InterPro; IPR001127; PTS\_EIIA.  
DR InterPro; IPR001996; PTS\_EIIB.  
DR InterPro; IPR003352; PTS\_EIIC.  
DR Pfam; PF00358; PTS\_EIIA\_1; 1.  
DR Pfam; PF00367; PTS\_EIIB; 1.  
DR Pfam; PF02378; PTS\_EIIC; 1.  
DR ProDom; PD002243; PTS\_EIIA; 1.  
DR ProDom; PD001476; PTS\_EIIB; 1.  
DR ProSITE; PS00371; PTS\_EIIA\_1; 1.  
DR ProSITE; PS10035; PTS\_EIIB\_CYS; 1.  
DR ProSITE; PS10035; PTS\_EIIB\_CYS; 1.  
KW Phosphotransferase system; Sugar transport; Transferase;  
KW Phosphorylation; Transmembrane; 3D-structure; Complete proteome.  
FT MOD\_RES 239 239 PHOSPHORYLATION (BY SIMILARITY).  
FT MOD\_RES 461 461 PHOSPHORYLATION (BY SIMILARITY).  
FT MOD\_RES 620 620 PHOSPHORYLATION (BY SIMILARITY).  
FT DOMAIN 1 ? EIIC DOMAIN.  
FT DOMAIN 1 ? EIIB DOMAIN.  
FT DOMAIN 530 690 EIIA DOMAIN.  
FT TRANSMEM 16 36 POTENTIAL.  
FT TRANSMEM 66 86 POTENTIAL.  
FT TRANSMEM 89 109 POTENTIAL.  
FT TRANSMEM 139 159 POTENTIAL.  
FT TRANSMEM 180 200 POTENTIAL.  
FT TRANSMEM 283 303 POTENTIAL.  
FT TRANSMEM 314 334 POTENTIAL.  
FT TRANSMEM 338 358 POTENTIAL.  
FT TRANSMEM 365 385 POTENTIAL.  
FT TRANSMEM 388 408 POTENTIAL.  
FT STRAND 544 544 POTENTIAL.  
FT TURN 546 547

FT STRAND 550 550  
FT STRAND 553 553  
FT STRAND 562 562  
FT HELIX 563 565  
FT HELIX 569 572  
FT TURN 573 574  
FT STRAND 578 584  
FT STRAND 588 590  
FT STRAND 595 600  
FT TURN 602 603  
FT STRAND 606 611  
FT TURN 612 613  
FT STRAND 616 620  
FT STRAND 623 623  
FT HELIX 626 628  
FT TURN 629 632  
FT STRAND 633 635  
FT TURN 639 640  
FT STRAND 642 643  
FT TURN 645 646  
FT STRAND 648 652  
FT HELIX 654 660  
FT STRAND 663 663  
FT STRAND 666 670  
FT TURN 675 676  
FT STRAND 679 681  
FT STRAND 686 687  
FT TURN 689 690  
FT STRAND 692 692  
FT STRAND 695 697  
SQ SEQUENCE 699 AA; 75525 MW; 2A14D3C32EE0A9C5 CRC64;  
  
Query Match 16.4%; Score 299; DB 1; Length 699;  
Best Local Similarity 24.3%; Pred. No. 5.9e-14;  
Matches 119; Conservative 55; Mismatches 134; Indels 182; Gaps 18;  
  
QY 7 VGGLLFGLVSPVITGLHQSF-PPIELELFN-----OGGSFIFATASMANIAQG- 55  
DB 221 LAAFIFGVIERSLIPGLHHIFYSFWEFFYSKAAAGEIIRGDDQRIF----MAQIKGV 276  
QY 56 -----AACLAFFELAKSEKLKLAGASGVA-----VLGITPAIF 91  
DB 277 QLTAGTFMTGKYPFMFGLPAAALAIYHEAKPQNKLVAGINGSRAALTSFLGITPELEF 336  
QY 92 G-----VNLRLRWPF-----I 103  
DB 337 SFLFVAPVLFALHCLFAGLSFPMVMOLLNVKIGMTFSGGLIDYFLFGILPNRTAWLVI 396  
QY 104 GIGTAAI--GGALIAL--FNK-----AVALGAAGFLGVVSIDAPDMVFLVC 149  
DB 397 GLGLAVIYFGFRFAIRKFNLTGPGREDAAEETAAPGKTGEAG---DLPYEIL----- 446  
QY 150 VTFFTAFG-----AAIAYGLVLRNRSIDPD----- 176  
DB 447 ----QAMGDQENIKHLDACIIRLVTVNDQKVKDRLKQLGASGVLEVGNIIQAFGPR 502  
QY 177 -----ATAAPVPAGTKAEAE-----PAEFSNDSTIIQAPLTGEATA 214  
DB 503 SDGLKTQMODIIAGRKPRPEKPSAQEEVGQVEEVIAEPLQNEIGEEVFVSPITGEIHP 562  
QY 215 LSSVSDAMFASGKLGSGVAIVPTKQLVSPVSGKIVVAFPPSGHAFVTRKADGNSVDIL 274  
DB 563 ITDVPDQVFSKMMGDGFALIPSEGVIVSPVRGKTLNVPPTKHAIGLOS---DGGR-EIL 618  
QY 275 MHIGFDTVNLNCTHENPLKKGDEDFDIDAKAAGYEVVTPPIVVS----- 328  
DB 619 IHFGIDTVSLKGEGETSFVSEGDVRPEGQKLLVDLDVAVKPNVPSLMTPIVFTNLAEGT 678  
QY 329 --YKKTGPVN 336  
DB 679 VSIKASGSVN 688

```
RESULT 14
PTSB_SALTY
ID PTB_SALTY STANDARD; PRT; 456 AA.
AC P08470.
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE PTS SYSTEM, SUCROSE-SPECIFIC IIBC COMPONENT (EIIBC-SCR) (SUCROSE-
DE PERMEASE IIBC COMPONENT) (PHOSPHOTRANSFERASE ENZYME II, BC COMPONENT)
DE (EC 2.7.1.69) (EII-SCR).
GN SCRA..
OS Salmonella typhimurium.
OG Plasmid pUR400.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88216186; PubMed=3285123;
RA Ebner R., Lengeler J.W.;
RT "DNA sequence of the gene scrA encoding the sucrose transport protein
RT EnzymeII(Scr) of the phosphotransferase system from enteric bacteria:
RT homology of the EnzymeII(Scr) and EnzymeII(Bgl) proteins.";
RL Mol. Microbiol. 2:9-17(1988).
RN [2]
RP REVISIONS.
RX MEDLINE=94018607; PubMed=8412665;
RA Jahreis K., Lengeler J.W.;
RT "Molecular analysis of two ScrR repressors and of a ScrR-FruR hybrid
RT repressor for sucrose and D-fructose specific regulons from enteric
RT bacteria.";
RL Mol. Microbiol. 9:195-209(1993).
RN [3]
RP SEQUENCE OF 1-7 FROM N.A.
RC STRAIN=6153-62;
RX MEDLINE=91100329; PubMed=1846143;
RA Hardesty C., Ferran C., Drenzo J.M.;
RT "Plasmid-mediated sucrose metabolism in Escherichia coli:
RT characterization of scrV, the structural gene for a
RT phosphoenolpyruvate-dependent sucrose phosphotransferase system
RT outer membrane porin.";
RL J. Bacteriol. 173:449-456(1991).
CC -!- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT
CC SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE
CC -TRANSPORT SYSTEM. THE IICD DOMAINS CONTAIN THE SUGAR BINDING SITE
CC AND THE TRANSMEMBRANE CHANNEL; THE IIA DOMAIN CONTAINS THE PRIMARY
CC PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HPR); IIA TRANSFERS ITS
CC THE SUGAR. EIIBC-SCR BELONGS TO THE EIIBC DOMAIN TYPE AND LACKS A
CC COVALENTLY BOUND EIITA DOMAIN. INSTEAD, EII-SCR-MEDIATED
CC PHOSPHORYLATION OF SUCROSE REQUIRES THE ACTIVITY OF ENZYME IIA-
CC GLC, A COMPONENT OF THE MAJOR GLUCOSE TRANSPORT SYSTEM.
CC -!- CATALYTIC ACTIVITY: PROTEIN N-PHOSPHOHISTIDINE + SUGAR =
CC PROTEIN HISTIDINE + SUGAR PHOSPHATE.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE.
CC -!- SIMILARITY: CONTAINS A PTS EIIIB DOMAIN.
CC -!- SIMILARITY: CONTAINS A PTS EIIBC DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X67750; CAA47973.1; .
DR EMBL; Y00541; CAA68605.1; ALT_SEQ.
DR EMBL; X38416; AAA98418.1; .
DR PIR; S01036; WQEST.
DR HSSP; P05053; IIBA.
DR InterPro; IPR001996; PTS_EIIB.
```

```
DR InterPro; IPR003352; PTS_EIIC.
DR Pfam; PF00367; PTS_EIIB; 1.
DR Pfam; PF02378; PTS_EIIC; 1.
DR ProDom; PD001476; PTS_EIIB; 1.
DR PROSITE; PS01035; PTS_EIIB_CYS; 1.
KW Phosphotransferase system; Transferase; Sugar transport; Plasmid;
KW Transmembrane; Inner membrane; Phosphorylation.
FT DOMAIN 1 ? EIIB DOMAIN.
FT MOD_RES 26 456 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 309 309 PHOSPHORYLATION (BY SIMILARITY).
SQ SEQUENCE 456 AA; 47864 MW; 98A6F1620AE50885 CRC64;

Query Match 16.3%; Score 296.5; DB 1; Length 456;
Best Local Similarity 40.3%; Pred No. 5.4e-14;
Matches 64; Conservative 29; Mismatches 63; Indels 3; Gaps 1;

QY 5 GPVGLLGLVYSPVITGLHQSFPPIELELFNQ---GGSFIFATASMANIAQGAACLA 61
DB 289 GWLAGLLFGLYSVIVITGIHHSFHAVEAGLLGNPSIGVNFLLPIWAMANVAQGGACLA 348
QY 62 FFLAKSEKLKLAGAGSVSAVLGITTEPAIFGVNLRWPFIFIGTAAIGGALIALFNK 121
DB 349 WFKTTDAKIKAITLPSAFSAMLGITTEAAIFGINLRVFKPIAALIGGAAGGAWVSVHY 408
QY 122 AVALGAAGFLGVVSDIDPDVMFMFLVCAVVTFFTAAGAAI 160
DB 409 MTAVGLTAIPGMAIVQASLLNVIIGMVIAGVAFVSL 447

RESULT 15
ID PTAA_ECOLI STANDARD; PRT; 648 AA.
AC P09323;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PTS SYSTEM, N-ACETYLGLUCOSAMINE-SPECIFIC IIBC COMPONENT (EIIBC-NAG)
DE (N-ACETYLGLUCOSAMINE-PERMEASE IIBC COMPONENT) (PHOSPHOTRANSFERASE
DE ENZYME II, ABC COMPONENT) (EC 2.7.1.69) (EII-NAG).
GN NAGE OR PSTN OR H0679.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88212176; PubMed=3284790;
RA Rogers M.J., Ohgi T., Plumbidge J., Soell D.;
RT "Nucleotide sequences of the Escherichia coli nage and nagB genes:
RT the structural genes for the N-acetylglucosamine transport protein of
RT the bacterial phosphoenolpyruvate: sugar phosphotransferase system
RT and for glucosamine-6-phosphate deaminase.";
RL Gene 62:197-207(1988).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=89050950; PubMed=3056518;
RA Peri K.G., Waygood E.B.;
RT "Sequence of cloned enzyme IIN-acetylglucosamine of the
RT phosphoenolpyruvate:N-acetylglucosamine phosphotransferase system of
RT Escherichia coli.";
RL Biochemistry 27:6054-6061(1988).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
```

[4]  
RP SEQUENCE FROM N.A.  
RC STRAIN-K12;  
RX MEDLINE=97061202; PubMed=8905232;  
RA Oshima T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A.,  
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,  
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,  
RA Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,  
RA Sanei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,  
RA Yano M., Horiuchi T.;  
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome  
corresponding to the 12.7-28.0 min region on the linkage map.";  
RL DNA RES. 3:137-155(1996).  
CC -!- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT  
CC SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE  
CC -TRANSPORT SYSTEM. THE IICD DOMAINS CONTAIN THE SUGAR BINDING SITE  
CC AND THE TRANSMEMBRANE CHANNEL; THE IIA DOMAIN CONTAINS THE PRIMARY  
CC PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HPR); IIA TRANSFERS ITS  
CC PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO  
CC THE SUGAR.  
CC -!- CATALYTIC ACTIVITY: PROTEIN N-PHOSPHOHISTIDINE + SUGAR =  
CC PROTEIN HISTIDINE + SUGAR PHOSPHATE.  
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE.  
CC -!- SIMILARITY: CONTAINS A PTS EIIA DOMAIN.  
CC -!- SIMILARITY: CONTAINS A PTS EIIB DOMAIN.  
CC -!- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.  
CC -----  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M19284; AAA24192.1; .  
DR EMBL; AE000171; AAC73773.1; .  
DR EMBL; D90706; BAA35322.1; .  
DR EMBL; D90707; BAA35327.1; .  
DR PIR; B29895; WQEC2N.  
DR PIR; A28896; A28896.  
DR HSSP; P08837; 2F3G.  
DR EcoGene; EG10635; nagE.  
DR InterPro; IPR001127; PTS\_EIIA.  
DR InterPro; IPR001996; PTS\_EIIB.  
DR InterPro; IPR003352; PTS\_EIIC.  
DR Pfam; PF00358; PTS\_EIIA\_1; 1.  
DR Pfam; PF00367; PTS\_EIIB; 1.  
DR Pfam; PF02378; PTS\_EIIC; 1.  
DR ProDom; PD001476; PTS\_EIIB; 1.  
DR ProDom; PD002243; PTS\_EIIA; 1.  
DR PROSITE; PS00371; PTS\_EIIA\_1; 1.  
DR PROSITE; PS01035; PTS\_EIIB\_CYS; 1.  
KW Phosphotransferase system; Sugar transport; Transferase;  
KW Phosphorylation; Transmembrane; Inner membrane; Complete proteome.  
FT DOMAIN 1 ?  
FT DOMAIN 393 427  
FT DOMAIN 517 648  
FT MOD\_RES 190 190  
FT MOD\_RES 412 412  
FT MOD\_RES 569 569  
FT TRANSMEM 16 36  
FT TRANSMEM 38 58  
FT TRANSMEM 70 90  
FT TRANSMEM 92 112  
FT TRANSMEM 132 152  
FT TRANSMEM 159 179  
FT TRANSMEM 192 212  
FT TRANSMEM 232 252  
FT TRANSMEM 260 280  
FT TRANSMEM 284 304  
FT TRANSMEM 308 328  
FT TRANSMEM 339 359

SQ SEQUENCE 648 AA; 68346 MW; 1E24C97CFCBAA59 CRC64;  
Query Match 16.2%; Score 295; DB 1; Length 648;  
Best Local Similarity 23.2%; Pred. No. 1e-13;  
Matches 114; Conservative 62; Mismatches 162; Indels 154; Gaps 12;  
QY 5 GPVGGLLFGLVSPYIVITGLHQSFPP-----EELFNQO---GSF 42  
DB 170 GALSGGIFGNRLIPLTGHQVLTNIAWFQIGFETNAAGTVPHGDINRYPACDGTAGMF 229  
QY 43 IFATASMANIAQAACLAFLVFLAKSEK---LKLGAAGSVSAVL-GITEPAIF----- 91  
DB 230 MSGFFPIMFGLPGAALAMYFAAPKERRPMVGMGLLSVAVTALTGTGTETLEFLFELAP 289  
QY 92 -----GVNLRURWPFPIGIGTAAIGGAL----- 114  
DB 290 LLYLLHALLTGISLGFATLLGIHAGFSFSAIDYALMYNLPAASQNVVWMLLMVGMVIFA 349  
QY 115 -----TALFNKA-----VALGAA----- 128  
DB 350 IYVVFSLVIRFMFLNLTPTGREKDEIVTEEANSNTEGLTQLATNYIAVAGTDMKAI 409  
QY 129 -----GFLGVVSIDAPDMVFLVCAVVTFFIAFGAAIAYG 163  
DB 410 DACITRLRLTVADRSARVNDTMCKRLGASGVVKNKQTIQIVGKAESIGDANKKVVARG 469  
QY 164 LYLVRNGSIDPDATAAPVPAGTTKAEAPAEFSNDSTIIQAPLTGEATALSVDAMF 223  
DB 470 -PVAASAEATP-ATAAPVAKPOAVPNVSIAB-----LVSPITGDVVVALDQVPDEAF 520  
QY 224 ASGKLGSVAIVPTKQLVSPVSGKIIVAFPPSGHAFVATKAEDGSNVLDLMHIGFDTVN 283  
DB 521 ASKAVGDGVAVRPTDKIVVSPAAGTIVKIFNTNHFACLETE-----KGAELVHMGIDITVA 576  
QY 284 LMGTHFNPLKKGDEVKAGELICEFDIDAIAKAGYEVTTPIVVSNTKKTGPVNTYGLGEI 343  
DB 577 LEGKGFRLVEEGAQVSAGQPILEMDLDVLNANARSMISPVVCSNIDDFSGLIKAQGI 636  
QY 344 EAGANLLNVAKK 355  
DB 637 VAGQTPLYEIKK 648  
Search completed: March 21, 2002, 16:28:51  
Job time: 331 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5

Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 21, 2002, 16:23:46 ; Search time 25.85 Seconds

(without alignments)

1066.738 Million cell updates/sec

Title: US-09-604-231-4

Perfect score: 1821

Sequence: 1 YDFGPGVGLLGLVSPIV.....IEAGANLLNVAKKAVPATP 362

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR\_68:\*

1: pir1.\*

2: pir2.\*

3: pir3.\*

4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query	Match	Length	ID	Description
1	711	39.0	664	1	B32243	phosphotransferase
2	706.5	38.8	651	2	S46257	phosphotransferase
3	562.5	30.9	630	2	H83686	PTS system, beta-g
4	550.5	30.2	636	2	C83724	PTS system, beta-g
5	515.5	28.3	632	2	S68599	phosphotransferase
6	515	28.3	609	2	I40406	beta-glucoside per
7	503	27.6	609	2	T47097	hypothetical prote
8	492.5	27.0	631	2	B42603	beta-glucoside-spe
9	450	24.7	625	2	C25977	phosphotransferase
10	448	24.6	636	2	D86807	hypothetical prote
11	446.5	24.5	480	2	S39978	sarA protein - Sta
12	416	22.8	479	2	F82432	PTS system, sucros
13	412	22.6	479	2	J00781	sucrose uptake pro
14	350.5	19.2	470	2	H83926	PTS system, trehal
15	330	18.1	470	2	C69725	phosphotransferase
16	314	17.2	195	2	D86766	hypothetical prote
17	313.5	17.2	692	1	S46953	phosphotransferase
18	313	17.2	651	1	S18607	phosphotransferase
19	310	17.0	458	2	H83881	PTS system, sucros
20	309	16.7	460	2	A39938	phosphotransferase
21	299	16.4	699	1	W0B5S	phosphotransferase
22	295	16.2	648	1	W0EC2N	phosphotransferase
23	294	16.1	648	2	E85567	hypothetical prote
24	293.5	16.1	456	2	S62331	phosphotransferase
25	290	15.9	372	2	I39868	sac operon regulat
26	277.5	15.2	455	1	W0EBST	phosphotransferase
27	276	15.2	675	2	D83755	PTS system, glucos
28	273	15.0	459	2	JU0293	levansucrase synth
29	271.5	14.9	173	2	C83839	PTS system, glucos

30 270.5 14.9 631 2 D69750 phosphotransferase  
31 269 14.8 675 1 S46952 phosphotransferase  
32 256 14.1 189 2 F70169 phosphotransferase  
33 255 14.0 168 2 H69940 phosphotransferase  
34 252.5 13.9 634 2 A32241 lactose transport  
35 249.5 13.7 473 2 C65236 phosphotransferase  
36 246 13.5 169 1 W0ECP3 phosphotransferase  
37 246 13.5 169 2 D85884 PTS system, glucos  
38 245 13.5 169 2 G82257 PTS system, glucos  
39 244 13.4 169 1 W0EB3T phosphotransferase  
40 243.5 13.4 478 2 B82263 PTS system, trehal  
41 242.5 13.3 473 2 A86122 PTS system enzyme  
42 238.5 13.1 166 2 G64137 phosphotransferase  
43 231.5 12.7 161 2 B84937 phosphotransferase  
44 226.5 12.4 842 2 C83177 probable phosphotr  
45 221 12.1 641 2 S44253 raffinose carrier

## ALIGNMENTS

RESULT 1

B32243

phosphotransferase system enzyme II (EC 2.7.1.69), sucrose-specific - Streptococcus m  
C:Species: Streptococcus mutans  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
C:Accession: B32243

R:Sato, Y.; Poy, F.; Jacobson, G.R.; Kuramitsu, H.K.

J. Bacteriol. 171, 263-271, 1989

A:Title: Characterization and sequence analysis of the scrA gene encoding enzyme II(s  
A:Reference number: A32243; MUID:89123027

A:Accession: B32243

A&gt;Status: preliminary

A:Molecule type: DNA

A:Residues: 1-664 &lt;SAT&gt;

A:Cross-references: GB:M22711; NID:g153799; PIDN:AAA26971.1; PID:g153801

C:Superfamily: phosphotransferase system enzyme II sucrose-specific; phosphotransfera

C:Keywords: phosphoprotein; phosphotransferase

F:511-664/Domain: phosphotransferase system glucose-specific enzyme II, factor III ho

Query Match 39.0%; Score 711; DB 1; Length 664;

Best Local Similarity 42.1%; Pred. No. 5.4e-43;

Matches 155; Conservative 73; Mismatches 116; Indels 24; Gaps 8;

Qy 1 YDFGPGVGLLGLVSPIVITGLHQSPPIEL- ---FNQG---GSFIFATASMANIAQ 54

Db 307 YDTTGLGCVGFGALYSPVMTGLHQSFPATETQLISAFQNGTGHGDFITFTASMANVAQ 366

Qy 55 GAACLAFFLAKSEKLGAGSVSAVLGITEPAIFGVNLRMRPFFIGITGAIGGAL 114

Db 367 GAATFAIYFLTKDKMKKGLSSSGVSALLGITEPAIFGVNLRMRPFFIGITGAIGGAL 426

Qy 115 IALNIKAVLGAAGFLGVSIDAPDMVFLVCVVTFIAPGAAIAYGLVLRNGSID 174

Db 427 AGLLQVAVSLGAGFLGLSLKASSIPFYVCELIISFAIAFAVITYG- ----KTKAVD 481

Qy 175 PDATAAPYPAGTTKAEAPAEFFSNDST- ----IIQAPLTGEAIALSSVSDAMFASGK 227

Db 482 VFAAEAAVEAEIEVQ- EIPPEAASAANKAQVTDVLAAPLAGEAVELTSVNDPVFSEA 540

Qy 228 LGSVAIVPTKQLVSPVSGKIVAFPSGHAFVATKADGDSNVLDILMHIGFDFTVNLNGT 287

Db 541 MGKGIAIKPSGNTVYAPVDGTVTAFTGCHAYGI- --KSDNGA- --BILIHIGIDTVSMGK 596

Qy 288 HFNPLKKGDEKAGELCEFDIDAIAKAAGVEVTPPIVVSNNKKTGPNTY- GLGEIAG 346

Db 597 GFEQKVQADQIKKGVDVLGTFDSKDIABAGLDNTTFTVNTADYASVETLASSGTVAVG 656

Qy 347 ANLLNVAK 354

Db 657 DSLLEVK 664

## RESULT 2

S44257  
phosphotransferase system enzyme II (EC 2.7.1.69), sucrose specific - Pedicoccus pentosae  
C:Species: pedicoccus pentosaceus  
C:Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 16-Jul-1999  
C:Accession: S44257  
R:Leenhouts, K.K.J.; Bolhuis, A.A.; Kok, J.J.; Venema, G.G.  
submitted to the EMBL Data Library, April 1994  
A:Description: The sucrose and raffinose operons of Pedicoccus pentosaceus PPE1.0.  
A:Reference number: S44252

A:Accession: S44257  
A:Molecule type: DNA  
A:Residues: 1-651 <LEE>  
A:Cross-references: EMBL:232771; NID:9493728; PIDN:CAA83668.1; PID:9475968  
C:Genetics:

A:Gene: scrA  
C:Superfamily: phosphotransferase system enzyme II sucrose-specific; phosphotransferase  
C:Keywords: phosphotransferase

F:488-640/Domain: phosphotransferase system glucose-specific enzyme II, factor III homol

Query Match 38.8%; Score 706.5; DB 2; Length 651;

Best Local Similarity 42.0%; Pred. No. 1.1e-42;  
Matches 154; Conservative 64; Mismatches 126; Indels 23; Gaps 5;

QY 1 YDFGGPVGGLLGLVYSPITVITGLHQSFPPIELF---NOGSGSFIFATASMANIAQGA 56

DB 300 YNSTGHWGMGIFGLLYSAIVITGLHQTFFPAETQLLANVAKTGSGSIFFPVSMANICQGA 359

QY 57 ACLAVFLLAKSEKLGKLAGAGSVAVLGIPTFAIFGVNLRRLRMPFFIGIGTAAIGGALIA 116

DB 360 ATLAIFGATKSKOKALTSSAGVSALLGITPAIFGVNLRMKRPFVFAALIASGASAFGL 419

QY 117 LFNKAVLGAAGFLGWSIDAPDMVFLVCAVVTFFIAFGAAIAYGLVLRNGSIDPD 176

DB 420 LFHVLVSMPASVIGFISIAKSIPAFMLSAVISVVAIFPTFIY---AKRTLGDGDRD 475

QY 177 ATAAPVPAGTTKAEAPAEFSNDSTIIQAPLTGEAIALSSVSDAMFASGKLGSGVAIVP 236

DB 476 QVKSAPTSTV-----INVNDRLISAPVTGASELSKVNDQVFSAEIMKGGAIVP 526

QY 237 TKQLVSPVSGKIVVAFPPSGHAFVTRKADGSDNVDILMHIGPDTVNLNGTHFNPLKKQG 296

DB 527 SSDQVVPADGVITVYDSSHAYGIKTTA---GAEILIHGLDVTNLNGEHEFTTNVQKG 582

QY 297 DEYKAGELLCEFDIDAIAKAAGYEVTTPIVVSNVKKTPVNTYGLGLEIEAGANL--LNVAK 354

DB 583 DTVHQDGLGTGFDIAALKAAVNDPTVMLIVNTANYANVERLKVTVNQAGEQLVALTAPA 642

QY 355 KEAVPAT 361

DB 643 ASSVAAT 649

## RESULT 3

H83686  
PTS system, beta-glucoside-specific enzyme II, ABC component BH0296 [imported] - Bacillus

C:Species: Bacillus halodurans  
C:Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 31-Dec-2000  
C:Accession: H83686

R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira  
Nucleic Acids Res. 28, 4317-4331, 2000

A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and  
A:Reference number: A83650; MUID:20263314

A:Accession: H83686  
A>Status: preliminary  
A:Molecule type: DNA

A:Residues: 1-630 <STO>  
A:Cross-references: GB:AP001508; GB:BA000004; NID:g10172890; PIDN:BAB04015.1; GSPDB:GN00

A:Experimental source: strain C-125  
A:Gene: BH0296

C:Superfamily: phosphotransferase system enzyme II sucrose-specific; phosphotransfera

Query Match 30.9%; Score 562.5; DB 2; Length 630;

Best Local Similarity 34.6%; Pred. No. 1.8e-32;  
Matches 124; Conservative 78; Mismatches 143; Indels 13; Gaps 4;

QY 1 YDFGGPVGGLLGLVYSPITVITGLHQSFPPIEL-ELFNOGSGSFIFATASMANIAQGAACL 59

DB 282 FTFSPLVLAGAIVAGIQVILVIFGIHGIPIIILNNLSVRGEDIKAVAPAVFSQAGAAL 341

QY 60 AVFFLAKSEKLGKLAGAGSVAVLGIPTFAIFGVNLRRLRMPFFIGIGTAAIGGALIALFN 119

DB 342 GVMRLTKNKKLALAGSTITALFGITEPAVYGVTLPLKKPFTMAVISAAGVGAIVGHY 401

QY 120 IKAVLGAAGFLGVSI---DAPDMVFLVCAVVTFFIAFGAAIAYGLVLRNGSIDPD 176

DB 402 SVAVAPGAPGLTIPTIFPEDGRGFVAFVIAIISVLA-----AVLTIVGVKDPVDDE 456

QY 177 ATAAPVPAGTTKAEAPAEFSNDSTIIQAPLTGEAIALSSVSDAMFASGKLGSGVAIVP 236

DB 457 DTLSNMGSENEVKREDDKKEPSASEIKSPLKGEVVPVLEVDHVFSSGAMGKGVAVRP 516

QY 237 TKQLVSPVSGKIVVAFPPSGHAFVTRKADGSDNVDILMHIGPDTVNLNGTHFNPLKKQG 296

DB 517 KEGRLVAPINGTVTSLFETKHAIGITS---DNGTEIFTHVIGIDTVOLKGEHFTSFIEQG 572

QY 297 DEYKAGELLCEFDIDAIAKAAGYEVTTPIVVSNVKKTPVNTYGLGLEIEAGANLNVAK 354

DB 573 DEVAAGDVLLEFDVERITAAGYDVIPTVLITNAKQFSNVOTTDKREVTSEDLIIHVIK 630

## RESULT 4

## C83724

PTS system, beta-glucoside-specific enzyme II, ABC component bglp [imported] - Bacillus

C:Species: Bacillus halodurans  
C:Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 08-Dec-2000  
C:Accession: C83724

R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira  
Nucleic Acids Res. 28, 4317-4331, 2000

A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a  
A:Reference number: A83650; MUID:20263314

A:Accession: C83724  
A>Status: preliminary  
A:Molecule type: DNA

A:Residues: 1-636 <STO>  
A:Cross-references: GB:AP001509; GB:BA000004; NID:g10173176; PIDN:BAB04314.1; GSPDB:G

A:Experimental source: strain C-125  
A:Gene: bglp  
C:Superfamily: phosphotransferase system enzyme II sucrose-specific; phosphotransfera

Query Match 30.2%; Score 550.5; DB 2; Length 636;

Best Local Similarity 36.7%; Pred. No. 1.3e-31;  
Matches 132; Conservative 55; Mismatches 150; Indels 23; Gaps 6;

QY 1 YDFGGPVGGLLGLVYSPITVITGLHQSFPPIEL-ELFNOGSGSFIFATASMANIAQGAACL 59

DB 294 YNLSPIIAGAFILGGFQVVFIFGLHWGLIPIAINNLVVOGSDPVLAMVFAASPAQIGAVA 353

QY 60 AVFFLAKSEKLGKLAGAGSVAVLGIPTFAIFGVNLRRLRMPFFIGIGTAAIGGALIALFN 119

DB 354 AWWLKIKQKQKVTLSVPFISGIFGVTEPAIYGVTLPLKRPFIISCIIAAAGVGAIGLFLR 413

QY 120 IKAVLGAAGFLGVSI---DAPDMVFLVCAVVTFFIAFGAAIAYGLVLRNGSID 174

DB 414 SQCYIIGLIGFIPSLHPADMDAGFWGIVAVVAVFV-LGFILTYLGLKSGNASDE 472

QY 175 PDATAPVPAGTTKAEAPAEFSNDSTIIQAPLTGEAIALSSVSDAMFASGKLGSGVAI 234

DB 473 QTEKTAHTSTGTGEKEE-----ISSPFGNSVITLSEIKRDEAFSSGALGEGIAI 520



QY 235 VPTKQLVSPVSGKIWAFFPSGHAFVRTKAEDGNSVDILMHIGFDTVNLNGTHFNPLKK 294  
Db 521 ESEKGLFSPVSGMTALYPHHLGITT-----DRGAELLIHIGLDTVOLDGKFFTAHTI 576  
QY 295 QGDEYKAGELCEFDIDAKAAGYEVVTPPIVVSNNKKTGPVNTYGLGETEAGANLNNAK 354  
Db 577 QGAQVEKGDLLIEFDIKEKAAGYATVPPIVITNKHQGLPLTDKQVNAAGDRLLLELR 636  
RESULT 5  
S85581  
phosphotransferase system enzyme II (EC 2.7.1.69), sucrose-specific - Streptococcus sobrius  
N:Alternate names: sucrose-specific enzyme II  
C:Species: Streptococcus sobrius  
A:Variety: strain 6715  
C:Date: 14-Feb-1997 #sequence\_revision 13-Mar-1997 #text\_change 07-May-1999  
C:Accession: S68599  
R:Chen, Y.Y.M.; Lee, L.N.; LeBlanc, D.J.  
Infect. Immun. 61, 2602-2610, 1993  
A:Title: Sequence analysis of scrA and scrB from Streptococcus sobrius 6715.  
A:Reference number: S68598; MUID:93273516  
A:Accession: S68599  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-632 <CHP>  
A:CROSS-references: EMBL:L06791  
C:Genetics:  
A:Gene: scrA  
C:Superfamily: phosphotransferase system enzyme II sucrose-specific; phosphotransferase  
C:Keywords: phosphotransferase; sugar transport system  
F:480-632/Domain: phosphotransferase system glucose-specific enzyme II, factor III homol

Query Match 28.3%; Score 515.5; DB 2; Length 632;  
Best Local Similarity 36.3%; Pred. No. 3.9e-29;  
Matches 128; Conservative 63; Mismatches 145; Indels 17; Gaps 7;  
QY 3 FGGPVGGLLFGVLVSPVITGLHQSFPPIELFNOGGSFIF-ATASMANIAQGAACLA 61  
Db 296 FG--TAGIIGGLQGVVTVGVVHHFNLETLQTLAETKANPFPNLLSAATAGVQAVLAV 353  
QY 62 FFLAKSEKLLAGAGSVAVIGTEPAIFGVNLRRLRPFFIGTIGTAAIGGALIALFNK 121  
Db 354 AVKTSKALKALYPSALSAAIGTEPAIFGVNLRRLRPFFVGLVGGAGGFIALLVGLK 413  
QY 122 AVALGAAGFLGVVSDADPMVMFLVCVVTFPIAAGAAIAYGLYLVRNGSIDPDATAP 181  
Db 414 ATGMSVTPLPLGLLFLNSQMPYIVSITVACAIAF--ALTY--YFGYADKEEDVSAKKPE 469  
QY 182 VPAGTTKAEAPAEFSNDSTIIQAPLTGEATLSSVSDAMFASGKLSGVAIVPTKGL 241  
Db 470 APAAPVPAETETKSE-----VIASPLDGEAVELSKVNDPVFSEAMGKIAVPSGNTV 523  
QY 242 VSPVSGKIWAFFPSGHAFVRTKAEDGNSVDILMHIGFDTVNLNGTHFNPLKKQGDVKA 301  
Db 524 YSPVNGTVQIAETGTHAGL--KSDNGA--EVLHVGIDTVSMNGTGFQDKVAANQTVK 579  
QY 302 GELLCEFDIDAKAAGYEVVTPPIVVSNNKKTGPVNTYGLGETEAGANLNNAK 354  
Db 580 GDVLGTFDSAKIAEAGLDDTTVIITNADYSEVVKPLAAGQLAHGADLLELNK 632  
RESULT 6  
I40406  
beta-glucoside permease - Bacillus subtilis  
C:Species: Bacillus subtilis  
C:Date: 12-Aug-1996 #sequence\_revision 12-Aug-1996 #text\_change 20-Jun-2000  
C:Accession: I40406; S65581; A69594; S47174  
R:Le Coq, D.; Lindner, C.; Kruger, S.; Steinmetz, M.; Stulke, J.  
J. Bacteriol. 177, 1527-1535, 1995  
A:Title: New beta-glucoside (bgl) genes in Bacillus subtilis: the bglp gene product has  
A:Reference number: I40404; MUID:95189730  
A:Accession: I40406

A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-609 <RES>  
A:CROSS-references: EMBL:Z34526; NID:G505573; PIDN:CAA84286.1; PID:G505576  
R:Beloin, C.; Hirschbein, L.; le Hegarat, F.  
Mol. Gen. Genet. 250, 761-766, 1996  
A:Title: Suppression of the Bgl+ phenotype of a Delta-hns strain of Escherichia coli  
A:Reference number: S65579; MUID:96204517  
A:Accession: S65581  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-74,'S',76-182 <BEL>  
A:CROSS-references: EMBL:X85408; NID:G1037169; PIDN:CAA59697.1; PID:G1037172  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, March 1995  
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, A.; Azevedo, V.; Ber  
C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;  
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,  
Nature 390, 249-256, 1997  
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal  
lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M  
Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino  
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau  
Y., M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete  
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanl  
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se  
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, K.; Tosato, V.; Uchiya  
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida  
A:Authors: Yoshikawa, H.F.; zumstein, E.; Yoshikawa, H.; Danchin, A.  
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis  
A:Reference number: A69580; MUID:98044033  
A:Accession: A69594  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-287,'F',289-433,'G',435,'S',437-448,'HR',451-548,'M',550-551,'S',553-60  
A:CROSS-references: GB:Z99124; GB:AL009126; NID:G2636442; PIDN:CAB15963.1; PID:G26364  
A:Experimental source: strain 168  
C:Genetics:  
A:Gene: bglp  
C:Superfamily: phosphotransferase system enzyme II sucrose-specific; phosphotransfera  
F:458-609/Domain: phosphotransferase system glucose-specific enzyme II, factor III ho

Query Match 28.3%; Score 515; DB 2; Length 609;  
Best Local Similarity 33.2%; Pred. No. 4e-29;  
Matches 121; Conservative 63; Mismatches 142; Indels 38; Gaps 6;  
QY 1 YDFGPGVGLLFGVLVSPVITGLHQSFPPIELFNOGGSFIFATASMANIAQGAAC 59  
Db 278 FDHAGLVAMILLAGTFLSIIMTMHYAFVPIINNTAQNQHDYLLPAMFLNMGQAGASF 337  
QY 60 AVFFLAKSEKLLAGAGSVAVIGTEPAIFGVNLRRLRPFFIGTIGTAAIGGALIALFN 119  
Db 338 AVFLSRNKKFKSLATTTISITLMTGITEPAMYGVNMLKKPFKFAAALIGGAAGAFYGMTG 397  
QY 120 IKAVALGA--AGFLGVVSDADPMVMFLVCVVTFPIAAGAAIAYGLYLVRNGSIDPDAT 178  
Db 398 VASYIVGGNAGLPSIPVFTGPTFIAMIGLVIAFAETAAAYLLGFEDVPDSGS----- 451  
QY 179 AAPVPAGTTKAEAPAEFSNDSTIIQAPLTGEATLSSVSDAMFASGKLSGVAIVPTK 238  
Db 452 -----QQPAVHEGSRRIIHSPIKGEVKSALSKVSGVAGVMGKGFALPEEE 498  
QY 239 GOLVSPVSGKIWAFFPSGHAFVRTKAEDGNSVDILMHIGFDTVNLNGTHFNPLKKQGD 298  
Db 499 GEWSVPVSGSVTTIFKTHAIGITS-----DQGAELIHIIGLDTVKLEGQWTAHIKEGDK 554  
QY 299 VKAGELCEFDIDAKAAGYEVVTPPIVVSNNKKTGPVNTYGLGETEAGANLNNAK 356  
Db 555 VAPGDLVPSFDLEQIKAAAGYDITPVITNTDOYSFSPVKEIG-----KVQPK 603  
QY 357 AVPA 360  
Db 604 ALLA 607

RESULT 7  
T47097  
hypothetical protein N17C [imported] - Bacillus subtilis  
C:Species: Bacillus subtilis  
C:Date: 17-Mar-2000 #sequence\_revision 17-Mar-2000 #text\_change 21-Jul-2000  
C:Accession: T47097  
R:Yoshida, K.; Sano, H.; Seki, S.; Oda, M.; Fujimura, M.; Fujita, Y.  
Microbiology 141, 337-343, 1995  
A:Title: Cloning and sequencing of a 29 kb region of the Bacillus subtilis genome contain  
A:Reference number: 224350; MUID:95219088  
A:Accession: T47097  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-609 <VOS>  
A:Cross-references: EMBL:D31856; NID:9603765; PIDN:BA00652.1; PID:9603778  
A:Experimental source: strain BGSC1A1  
C:Genetics:  
C:Superfamily: phosphotransferase system enzyme II sucrose-specific; phosphotransferase  
Query Match 27.6%; Score 503; DB 2; Length 609;  
Best Local Similarity 33.5%; Pred. No. 2.9e-28;  
Matches 122; Conservative 65; Mismatches 139; Indels 38; Gaps 9;  
QY 1 YDFGGPVGGLLFGLVYSPVITGLHOSFPPIELELEFNQGG-SPIFATASMANIAOQAACL 59  
DB 278 FDHAGLVAMIFLAGTSLIIMTGMHFAFVPMINNTAONGHYLLPAMFLANMQAGASF 337  
QY 60 AVFLLAKSEKLGLAGAGSVAVLGIPTAIFGVNLRWRPFFIGTAAIGGALIALFN 119  
DB 338 AVFLRGRNKKFKSLALTITLALMGITPEAMYGNNRMLKKPFAAALIGGAAGGAEYGMTG 397  
QY 120 IKAVALGA-AGFLGVVSIDAPDMVFLVCVAVTFFTAAGAAIAYGLYLVRNGSIDPDA 178  
DB 398 VASYIYVGNAGLPSIPVFITGPTFIYAMIGLVIAF--AAGTSAAYLL-----GFED---- 445  
QY 179 AADVPAGTTKAAEAAPAEFSNDSTIIQAPLTGEAIALSSVSDAMFASGKLGSVAIVPTK 238  
DB 446 ---VPS---HRSQQPAVHEGSRREIHSPIKGEVKALSEVKGVFSAGVMGKGFALEPEE 498  
QY 239 GQLVSPVSGKIVVAPSPGSHAFVTRTKAEDGNSVDILMHGIFDFTVNLNGTHFNPLKQGD 298  
DB 499 GEVVPVRSVTTIFTKKAIGITS----DOGAELIHIGLDIVKLEQWFTAHMKESDK 554  
QY 299 VKAGELLCFEDIDAIAKAGYEVTTPIVVSNYK--TGPVNTYGLGIEAGANLLNVAKKE 356  
DB 555 VAFGDPVLVSFDLEQIKAAAGYDVITPVVNTDQYSPSPVKEIG-----KVQPK 603  
QY 357 AVPA 360  
DB 604 ALIA 607  
RESULT 8  
B42603  
beta-glucoside-specific phosphotransferase system-dependent permease - Erwinia chrysanth  
C:Species: Erwinia chrysanthemi  
C:Date: 04-Sep-1998 #sequence\_revision 04-Sep-1998 #text\_change 16-Jul-1999  
C:Accession: B42603  
R:el Hassouni, M.; Henrissat, B.; Chippaux, M.; Barras, F.  
J. Bacteriol. 174, 765-777, 1992  
A:Title: Nucleotide sequences of the arb genes, which control beta-glucoside utilization  
of the family including enzymes from eubacteria, archaeobacteria, and humans.  
A:Reference number: A42603; MUID:92121114  
A:Accession: B42603  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-631 <ELI>  
A:Cross-references: GB:M81772; NID:9148385; PIDN:AAA24814.1; PID:9148387  
A:Note: sequence extracted from NCBI backbone (NCBIN:77535, NCBI:77537)

C:Genetics:  
A:Gene: arbF  
C:Superfamily: phosphotransferase system enzyme II sucrose-specific; phosphotransfera  
F:479-631/Domain: phosphotransferase system glucose-specific enzyme II, factor III ho  
Query Match 27.0%; Score 492.5; DB 2; Length 631;  
Best Local Similarity 34.9%; Pred. No. 1.7e-27;  
Matches 123; Conservative 62; Mismatches 154; Indels 13; Gaps 7;  
QY 1 YDFGGPVGGLLFGLVYSPVITGLHOSFPPIELELEFNQ--GGSFATASMANIAOQAACL 59  
DB 285 YGLNSLLAGAVMGALWQCVIFGLHWGFVPLMLNNSFVIGHDTLLPLLLPAVLQAGATL 344  
QY 60 AVFLLAKSEKLGLAGAGSVAVLGIPTAIFGVNLRWRPFFIGTAAIGGALIALFN 119  
DB 345 GVLLRTQDLKRKGAGSASFAIFGITEPAVYGVTTPLRRPFFIGGIGGALGAAMVGYAH 404  
QY 120 IKAVALGAAGFLGVVSIDAPDMVFLV-CAVYVTFIAFG-AAIAYGLYLVRNGSIDPDA 177  
DB 405 TTMYSFGFPSTFTQVIPPPTGVDSVWAAVIGTLLAFAFAALTSWSFGVPKD---ETQP 461  
QY 178 TAAPVPAGTTKAAEAAPAEFSNDSTIIQAPLTGEAIALSSVSDAMFASGKLGSVAIVPT 237  
DB 462 AAADSPA--VLAETQANAGAVRDETILF-SPLAGEVLLLEBOVADRTFASGVMGKGIARPT 518  
QY 238 GQLVSPVSGKIVVAPSPGSHAFVTRTKAEDGNSVDILMHGIFDFTVNLNGTHFNPLKQGD 297  
DB 519 OGELYAPVDGTVASLFKTHAIGLASR---GGAEVLIHWGIDTVRLDGRYFTPHVRVGD 574  
QY 298 EVKAGELLCFEDIDAIAKAGYEVTTPIVVSNYKKTGPVNTYGLGIEAGANL 349  
DB 575 VVRQGLLLEFDFGPAIEAAGYDLTTPITVNTSDEYRGVPEVSGVKVDANAPL 626  
RESULT 9  
C25977  
phosphotransferase system enzyme II (EC 2.7.1.69), beta-glucoside-specific - Escheric  
C:Species: Escherichia coli  
C:Date: 15-Dec-1988 #sequence\_revision 15-Dec-1988 #text\_change 18-Jun-1999  
C:Accession: C25977; A47616; C65175  
R:Schmetz, K.; Toloczky, C.; Rak, B.  
J. Bacteriol. 169, 2579-2590, 1987  
A:Title: Beta-glucoside (bgl) operon of Escherichia coli K-12: nucleotide sequence, g  
S.  
A:Reference number: A91833; MUID:87222180  
A:Accession: C25977  
A:Molecule type: DNA  
A:Residues: 1-625 <SCH>  
A:Cross-references: GB:M16487; GB:M60288; NID:g145414; PIDN:AAA23510.1; PID:g145418  
R:Bramley, H.F.; Kornberg, H.L.  
J. Gen. Microbiol. 133, 563-573, 1987  
A:Title: Nucleotide sequence of bglC, the gene specifying enzymeII(bgl) of the PEP:su  
A:Reference number: A47616; MUID:88009877  
A:Accession: A47616  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-625 <BRA>  
A:Cross-references: GB:M15746; NID:g145412; PIDN:AAA83837.1; PID:g145413  
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;  
A.; Rose, D.J.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997  
A:Title: The complete genome sequence of Escherichia coli K-12.  
A:Reference number: A64720; MUID:97426617  
A:Accession: C65175  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-625 <BLAT>  
A:Cross-references: GB:AE000449; GB:U00096; NID:g2367269; PIDN:AAC76745.1; PID:g17901  
A:Experimental source: strain K-12, substrain MG1655  
C:Genetics:  
A:Gene: bglF, bglS  
C:Superfamily: phosphotransferase system enzyme II sucrose-specific; phosphotransfera

C:Keywords: phosphoprotein; phosphotransferase; transmembrane protein  
F:473-625/Domain: phosphotransferase system glucose-specific enzyme II, factor III homol

Query Match 24.7%; Score 450; DB 2; Length 625;  
Best Local Similarity 33.3%; Pred. No. 1.7e-24;  
Matches 116; Conservative 64; Mismatches 154; Indels 14; Gaps 6;

QY 9 GLLFGLVSPVITGLHQSPPIELFNF-OGSFFIFATASMANIAQGAACLAFFLAKS 67

DB 290 GAVMGFWQIFVFMGLHGLVPLCLINNFVTLGYDTHPIPLMPALMAQVGAALGVFLCERD 349

QY 68 EKLKGLAGASGVAVLIGTEPAIFGVNLRWPFPIGCTAAGGALIALFNKKAVALGA 127

DB 350 AQKVVAGSAALSLFTEITPAVGVNLRPKYFVFIACISGALGATIIGYATQKVVSFGL 409

QY 128 AGFLGVVSDADPMVMFLVCAVVTFFIAFGAIAIYGLYLVRNNGSIDDP-ATAAPVPAGT 186

DB 410 PSIFTEMQITPSTGIDFTWASV-----IGGVIAIGCAFV--GTVMLHFIITAKRQPAQ 461

QY 187 TKAEAPAEFSNDSTIIQAPLTGEAIALSSVSDAMFASGKLGSGVAIVPTKGQLVSPVS 246

DB 462 APOEKTPEVITPEQGGICSPMTGEIVPLIHVADTTFASGLLGKGTAILPSVGEVRSVA 521

QY 247 GTIVVAFPSGHAFVTRKAEDGSNDVILMHIGRDTVNLNGTHFNPLKKGDEVKAGELLC 306

DB 522 GRISLFAFLHAIGI--ESDGG--VEILHVGIDTVKLDGKFSAHVNVGDKNVTGDRLI 577

QY 307 EFDIDAIAAGVEVTPPIVSVNKKTPVNTYGLGRIEAGANLLNVAK 354

DB 578 SFDIPAIREAGFDLTPVLISNDDFTDVLPHGTAISAGEPLLSIIR 625

RESULT 10

D86807  
hypothetical protein ptbA [imported] - Lactococcus lactis subsp. lactis (strain IL1403)

C:Species: Lactococcus lactis subsp. lactis  
C:Date: 23-Mar-2001 #sequence\_revision 23-Mar-2001 #text\_change 31-Mar-2001

C:Accession: D86807  
R:Botolin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarne, K.; Weissenbach, J.; Ehrlich

Genome Res. in press, 2001  
A:Title: The complete genome sequence of the lactic acid bacterium.

A:Reference number: A86625  
A:Accession: D86807

A>Status: preliminary  
A:Molecule type: DNA

A:Residues: 1-636 <STO>  
A:Cross-references: GB:AE005176; NID:g12724452; PTDN:AAK05558.1; GSPDB:GN00146

A:Experimental source: strain IL1403  
C:Genetics:

A:Gene: ptbA  
C:Superfamily: phosphotransferase system enzyme II sucrose-specific; phosphotransferase

Query Match 24.6%; Score 448; DB 2; Length 636;  
Best Local Similarity 33.0%; Pred. No. 2.4e-24;  
Matches 110; Conservative 59; Mismatches 136; Indels 28; Gaps 7;

QY 3 FCGPVGGLLGLVSPVITGLHQSPPIELFNOG-GSFIFATASMANIAQGAACLA 61

DB 288 FNPILFGLVGFPAQVQLVFMGLHVAIVPAITAIALAKGEPTALLIAASVASFAQTGAVGAV 347

QY 62 FFLAKEKLGKLAGASGVAVLIGTEPAIFGVNLRWPFPIGCTAAGGALIALFNK 121

DB 348 MLTKDKRLRELAIPAIFASGFWGTEPAIYGIITLPKRPFVWVSCIVSGILSAVAMALGIK 407

QY 122 AVAALGAGFLGVVSDADPMVMFLVCAVVTFFIAFGAIAIYGL--YLVRNNGSIDPDAT 178

DB 408 AYTWGA---IGIFSFTSNTLTGVEVSGAKIMIVSAVAVIAGFVTVLV---GFEDDVI 460

QY 179 AAPVPAGTTKAEEAPAEFSNDST---IIQAPLTGEAIALSSVSDAMFASGKLGSGVAV 235

DB 461 ENPIP-----DKKFNKOKTNKEIGSPLEGVIPLSQVKDAFASAGVMGKGAAT 510

QY 236 PTKGQLVSPVSKIVVAFPSGHAFVTRKAEDGSNDVILMHIGRDTVNLNGTHFNPLKKQ 295

DB 511 PTLGEVRAFPDGMVMILFPTKHAVALIS----NEGTELLIHIGIDTVQLEGKYFETFKVO 566

QY 296 GDEVRAGELLCEFDIDAIAKAAGVEVTPPIVWSN 328

DB 567 GQSVKGDILLKFDIERIQNAGYSTOVPIIVTN 599

RESULT 11

S39978

scrA protein - Staphylococcus xylosus

C:Species: Staphylococcus xylosus

C:Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 18-Jun-1999

C:Accession: S39978

R:Wagner, E.; Goetz, F.; Brueckner, R.  
Mol. Gen. Genet. 241, 33-41, 1993

A:Title: Cloning and characterization of the scrA gene encoding the sucrose-specific

A:Reference number: S39978; MUID:94049686

A:Accession: S39978

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-480 <WAG>

A:Cross-references: EMBL:X69800; NID:g407905; PIDN:CAA49461.1; PID:g407908

C:Superfamily: phosphotransferase system sucrose-specific enzyme II, factor II

Query Match 24.5%; Score 446.5; DB 2; Length 480;  
Best Local Similarity 52.5%; Pred. No. 2.2e-24;  
Matches 93; Conservative 30; Mismatches 47; Indels 7; Gaps 3;

QY 1 YDFGPGVGLLGLVSPVITGLHQSPPIELF-----NQGSFIFATASMANIAQGA 56

DB 301 YFEGGAIGLGLVGLVAPVITGMHSHFIAETQLIADSSSTGSGSFFIPFIATMSNIAQGA 360

QY 57 ACLAVFELAK-SEKLGKLAGASGVAVLIGTEPAIFGVNLRWPFPIGCTAAGGALI 115

DB 361 ALAALAFIIEKNKLGKVASAGVSALLGITEPAMFGVNLKLRYPFIGAIVSGSIGSAYI 420

QY 116 ALFNKAVAGAGFLGVVVSIDAPD--VMVFLVCAVVTFFIAFGAIAIYGLYLVRN 170

DB 421 APFKVKAIALGTAGIPGFTISIQNGNGLHYGIAMIIAFTVAFGVTVALSRYKKRN 477

RESULT 12

F82432

PTS system, sucrose-specific IIBC component VCA0653 [imported] - Vibrio cholerae (str

C:Species: Vibrio cholerae

C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001

C:Accession: F82432

R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.

chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers

l, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.

Nature 406, 477-483, 2000

A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.

A:Reference number: AB2035; MUID:20406833

A:Accession: F82432

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-479 <HEI>

A:Cross-references: GB:AE004395; GB:AE003853; NID:g9658068; PIDN:AAF96554.1; GSPDB:GN

A:Experimental source: serogroup O1; strain N16961; biotype El Tor

C:Genetics:

A:Gene: VCA0653

A:Map position: 2

C:Superfamily: phosphotransferase system sucrose-specific enzyme II, factor II

Query Match 22.8%; Score 416; DB 2; Length 479;  
Best Local Similarity 52.4%; Pred. No. 3.2e-22;  
Matches 86; Conservative 26; Mismatches 48; Indels 4; Gaps 1;



Query Match 18.1%; Score 330; DB 2; Length 470;  
Best Local Similarity 41.6%; Pred. No. 4e-16;  
Matches 67; Conservative 34; Mismatches 58; Indels 2; Gaps 1;

QY	7	VGGLFGLVYSPVITGLHQSFPPIELEFNQ--GGSFIFATASMANIAOGACLAVFEL	64
Db	307	LGGLLYGGFYSALVITGMHHTFLAVDLQLIGSKLGTFLWPMIALSNIAOGSAAIAMMFI	366
QY	65	AKSEKLGLAGAGSVSAVLGITEPAIFGVNLRWRPFFIGTAAIGGALIALFNKAVA	124
Db	367	VKDEKOKGLSLTSGISAYLGITPAIFGVNLRVRPFFIIAMVSSGLAGMYISSQGV	426
QY	125	LGAAGFLGVVSDAPDMVWFLVCANVTFFIAFGAAIAYGLY	165
Db	427	VGVGGVPGIFSIMSQYWGAFAGMAIVLIVPFGTYAYARF	467

Search completed: March 21, 2002, 16:23:47  
Job time: 153 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 21, 2002, 16:27:36 ; Search time 29.76 Seconds  
(without alignments)  
821.709 Million cell updates/sec

Title: US-09-604-231-4  
Perfect score: 1821  
Sequence: 1 YDFGPGVGLLGLVSPIV.....IEAGALLNVAKEAVPATP 362

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 322656 seqs, 67552660 residues

Total number of hits satisfying chosen parameters: 322656

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Pending\_Patents\_AA\_New.\*

- 1: /cgn2\_6/ptodata/1/paa/pct\_NEW\_COMB.pep.\*
- 2: /cgn2\_6/ptodata/1/paa/US05\_NEW\_COMB.pep.\*
- 3: /cgn2\_6/ptodata/1/paa/US07\_NEW\_COMB.pep.\*
- 4: /cgn2\_6/ptodata/1/paa/US08\_NEW\_COMB.pep.\*
- 5: /cgn2\_6/ptodata/1/paa/US09\_NEW\_COMB.pep.\*
- 6: /cgn2\_6/ptodata/1/paa/US09\_NEW\_COMB.pep1.\*
- 7: /cgn2\_6/ptodata/1/paa/US10\_NEW\_COMB.pep.\*
- 8: /cgn2\_6/ptodata/1/paa/US60\_NEW\_COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	542	29.8	627	1	PCT-US02-03987-13467
2	542	29.8	627	6	US-09-815-242-13467
3	542	29.8	627	7	US-10-072-851-13467
4	310.5	17.1	679	1	PCT-US02-03987-5658
5	310.5	17.1	679	6	US-09-815-242-5658
6	310.5	17.1	679	7	US-10-072-851-5658
7	310.5	17.1	681	1	PCT-US02-03987-12270
8	310.5	17.1	681	6	US-09-815-242-12270
9	310.5	17.1	681	7	US-10-072-851-12270
10	295	16.2	648	1	PCT-US02-03987-10089
11	295	16.2	648	6	US-09-815-242-10089
12	295	16.2	648	7	US-10-072-851-10089
13	276.5	15.2	473	6	US-09-897-516-6783
14	276.5	15.2	551	1	PCT-US02-03987-10500
15	276.5	15.2	551	6	US-09-815-242-10500
16	276.5	15.2	551	7	US-10-072-851-10500
17	274	15.0	687	1	PCT-US02-03987-5822
18	274	15.0	687	6	US-09-815-242-5822
19	274	15.0	687	7	US-10-072-851-5822
20	274	15.0	719	1	PCT-US02-03987-12942
21	274	15.0	719	6	US-09-815-242-12942
22	274	15.0	719	7	US-10-072-851-12942
23	247.5	13.6	169	6	US-09-897-516-6596
24	229	12.6	280	1	PCT-US02-03987-12479
25	229	12.6	280	6	US-09-815-242-12479

ALIGNMENTS

RESULT 1  
PCT-US02-03987-13467  
; Sequence 13467, Application PC/TUS0203987  
; GENERAL INFORMATION:  
; APPLICANT: Elitra Pharmaceuticals, Inc.  
; TITLE OF INVENTION: Methods for Identifying the Target of a Compound which Inhibit  
; TITLE OF INVENTION: Proliferation  
; FILE REFERENCE: ELTRA.028VPC  
; CURRENT APPLICATION NUMBER: PCT/US02/03987  
; PRIOR FILING DATE: 2002-02-02  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09  
; NUMBER OF SEQ ID NOS: 15811  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13467  
; LENGTH: 627  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
PCT-US02-03987-13467

Query Match 29.8%; Score 542; DB 1; Length 627;  
Best Local Similarity 36.0%; Pred. No. 1.3e-32;  
Matches 128; Conservative 74; Mismatches 130; Indels 24; Gaps 9;

QY	3	FGPGVGLLGLVSPDIVTGLHQSPPTLELFGNOGGSFIF-ATASMANIAOGAACLAV	61
DB	288	FG-LGGFTIGGVHQLIVYSGVHHIENLLEVLQAAADHANPENAIITAAATAOGAATVAV	345
QY	62	FLAKSEKLGLAGASGSAVLGITEPAIFGVNLRURWPFIFIGTAAIGGALIALFNK	121
DB	346	GVKTNPKUKTLAFPAALSAFLGITEPAIFGVNLRKPFELSLIAGAGGLASIGLA	405
QY	122	AVALGAAGFLG-VVSDADPMVNFVCAVVFTEFAFGAAIAYGLVLRNRSIDPDATAA	180
DB	406	GTGNGITITPTMLYVGNQLPQYLLMVAVSFALGALTMYFCY-----EDEVATAA	458
QY	181	PVPAGTKRAAE-APAEFSNDSTIIQAPLTGEAIALSSVSDAMFASGKLGSGVAIVPTRG	239
DB	459	AKQAEVAEKEEVAAPALQNETLV--TPIVGDVVALADVNDPVFSSGAMGQGIAPKPSQG	516
QY	240	QLVSPVSGKIVAVFSGHAFVTRKEDSNVDILMHIGFDIVNLGHTFNPPLKKGDEV	299
DB	517	VYVALADAEVSFAFTPTGHAFGLKTR---NGAEVLHVIGDITVSMNGDGFPEAKVAOGNKV	572
QY	300	KAGELLCEDDIAKAAAGVETTPVIWSN---YKKTGPVNTYGLGIEAGANLLN	352
DB	573	KAGDVLGTFDSNKIAAAGLUDDTTMTVITNTADYASVAPVAT---GSVSKGDAVIEV	625

Sequence 12479, A  
Sequence 11950, A  
Sequence 11950, A  
Sequence 11950, A  
Sequence 13018, A  
Sequence 13152, A  
Sequence 13018, A  
Sequence 13152, A  
Sequence 13018, A  
Sequence 13152, A  
Sequence 13018, A  
Sequence 13152, A  
Sequence 5842, Ap  
Sequence 5842, Ap  
Sequence 5842, Ap  
Sequence 10809, A  
Sequence 10809, A  
Sequence 10809, A  
Sequence 5718, Ap  
Sequence 5718, Ap  
Sequence 5718, Ap  
Sequence 12272, A





; NUMBER OF SEQ ID NOS: 15811
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5658
; LENGTH: 679
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
PCT-US02-03987-5658

Query Match 17.1%; Score 310.5; DB 1; Length 679;
Best Local Similarity 24.6%; Pred. No. 2.6e-15;
Matches 111; Conservative 54; Mismatches 149; Indels 137; Gaps 11;

QY 7 VGGLLFGLVYSPVITGLHQSF-----PPIELELFNOG----- 39
DB 209 VAVFLFGFKRLIIPGLHHIFHAPFWFEGSNKNAAGEIIHGDQRIEIQREGAHLTA 268
QY 40 GSIFATASMANIAQGAACLAFFLAKSEKLGLAGASGVA-----VLGITEPAIFGVNL 95
DB 269 GKFMQGEFPVMFGLPAAALAIYHSAPENKVVAGLMGSAALTSFLTGTITEPLEFSFLF 328
QY 96 RLRFWFFIGITAAAGGALIALENIKAVAG---AAGFLGVVSIDA-PDMVMP----- 144
DB 329 VAPLLFFI---HAVLDGLSFLTLVLDVHLGYTFSGGFDYVLLGVLPNKTQMWLVIPVG 385
QY 145 LVCVVTFPI-----AFGAIAAYGLY----- 165
DB 386 LVYAVIYFVFRFLIVKLKYKTPGREDKOSQAVTASATELPYAVLEAMGGKANIKHLDAC 445
QY 166 LVR-----RNGSIDPD 176
DB 446 ITRLREVNDSKVDVPGKLDLGASGVLEVGNMQAIFGPKSDQIKHEMQQIMNGQVVEN 505
QY 177 ATAAPVPAGTTKAEAPAEFNSDSTIIQAPLTGEATLSSVSDAMFASGKLGSGVAIVP 236
DB 506 PTTMEDDKDQETVVVAEDKTSATSELSHIVHAPLTGEVTPLEVPDQVFSEKMMGDGIAIKP 565
QY 237 TKGOLVSPVSGKIVVAFPSGHAFVTRKAEDSGNVDILMHIGFDTVNLNGTHFNPLKKOG 296
DB 566 SQGEVRAPNGKVMQIFPTKHAIGL-----VSDSGLELLIHIGLDTVKLNGEGFTLHVEEG 621
QY 297 DEVKAGELCEFDIDAIAKAAGYEVTTPIVVS 327
DB 622 QEVKQDGLLINFDLVDYIRNHAKSITPIIYT 652

RESULT 5
US-09-815-242-5658
; Sequence 5658, Application US/09815242
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5658
; LENGTH: 679
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5658

Query Match 17.1%; Score 310.5; DB 6; Length 679;
Best Local Similarity 24.6%; Pred. No. 2.6e-15;
Matches 111; Conservative 54; Mismatches 149; Indels 137; Gaps 11;

QY 7 VGGLLFGLVYSPVITGLHQSF-----PPIELELFNOG----- 39
DB 209 VAVFLFGFKRLIIPGLHHIFHAPFWFEGSNKNAAGEIIHGDQRIEIQREGAHLTA 268
QY 40 GSIFATASMANIAQGAACLAFFLAKSEKLGLAGASGVA-----VLGITEPAIFGVNL 95
DB 269 GKFMQGEFPVMFGLPAAALAIYHSAPENKVVAGLMGSAALTSFLTGTITEPLEFSFLF 328
QY 96 RLRFWFFIGITAAAGGALIALENIKAVAG---AAGFLGVVSIDA-PDMVMP----- 144
DB 329 VAPLLFFI---HAVLDGLSFLTLVLDVHLGYTFSGGFDYVLLGVLPNKTQMWLVIPVG 385
QY 145 LVCVVTFPI-----AFGAIAAYGLY----- 165
DB 386 LVYAVIYFVFRFLIVKLKYKTPGREDKOSQAVTASATELPYAVLEAMGGKANIKHLDAC 445
QY 166 LVR-----RNGSIDPD 176
DB 446 ITRLREVNDSKVDVPGKLDLGASGVLEVGNMQAIFGPKSDQIKHEMQQIMNGQVVEN 505
QY 177 ATAAPVPAGTTKAEAPAEFNSDSTIIQAPLTGEATLSSVSDAMFASGKLGSGVAIVP 236
DB 506 PTTMEDDKDQETVVVAEDKTSATSELSHIVHAPLTGEVTPLEVPDQVFSEKMMGDGIAIKP 565
QY 237 TKGOLVSPVSGKIVVAFPSGHAFVTRKAEDSGNVDILMHIGFDTVNLNGTHFNPLKKOG 296
DB 566 SQGEVRAPNGKVMQIFPTKHAIGL-----VSDSGLELLIHIGLDTVKLNGEGFTLHVEEG 621
QY 297 DEVKAGELCEFDIDAIAKAAGYEVTTPIVVS 327
DB 622 QEVKQDGLLINFDLVDYIRNHAKSITPIIYT 652

RESULT 6
US-10-072-851-5658
; Sequence 5658, Application US/10072851
; GENERAL INFORMATION:
; APPLICANT: Carr, Grant J.
; APPLICANT: Xu, H. Howard
; APPLICANT: Foulkes, J. Gordon
; APPLICANT: Zamudio, Carlos
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Roemer, Terry
; APPLICANT: Jiang, Bo
; APPLICANT: Boone, Charles
; APPLICANT: Bussey, Howard
; TITLE OF INVENTION: Methods for Identifying the Target of a Compound which Inhibit
; FILE REFERENCE: ELITRA.028A
; CURRENT APPLICATION NUMBER: US/10/072,851
; CURRENT FILING DATE: 2002-02-08

;; PRIOR APPLICATION NUMBER: 60/267,636  
;; PRIOR FILING DATE: 2001-02-09  
;; NUMBER OF SEQ ID NOS: 15811  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO: 5658  
;; LENGTH: 679  
;; TYPE: PRT  
;; ORGANISM: Staphylococcus aureus  
US-10-072-851-5658

Query Match 17.1%; Score 310.5; DB 7; Length 679;  
Best Local Similarity 24.6%; Pred. No. 2.6e-15;  
Matches 111; Conservative 54; Mismatches 149; Indels 137; Gaps 11;

```
QY 7 VGLLGLVYSPVITGLHOSF-----PPIELELFNOG----- 39
Db 209 VAVFLGFKRLIPFGLHHIFHAPFWFEGSNKNAAGEIHHGDQRIEIQREGAHLTA 268
QY 40 GSFIATASMANIAQAACLAFLAKSEKLGLAGASGVA-----VLGITEPAIFGVNL 95
Db 269 GKFMQGEFFVMFGLPAAALAIYHSAPENKVVVAGLMSAALTSTLTGITEPLEFSFLF 328
QY 96 RLKWPFFIGTGAAGGALIALFNKAVLG-----AAGFLGVVSDA-PDMVME----- 144
Db 329 VAPLLFFI-----HAVLDGLSFLTLVLLDVLHGYTFSGGFDYVLLGLVLPNKQTQWMLVIPVG 385
QY 145 LVCVVVTFE-----AFCAAIAYGLY----- 165
Db 386 LVAVIYVFFRFLVVKYKTPGREDKQSOAVTASATELPYAVLEAMGKGNKIKHLDAC 445
QY 166 LVR-----RNGSIDPD 176
Db 446 ITRLRVEVNDKSKVDVPLKDLGASGVLEGVNNMQAIFGPKSDQIKHEMQOIMNGQVVEN 505
QY 177 ATAAPVPAGTTKAEAPAEFSDNSTIIQAPLTGEATLSSVSDAMPASGKLGSGVAIVP 236
Db 506 PTMEDDKDETIVVVAEDKTSATSELSHIVHAPLTGEVTPLEVPDQVFESEKMMGSGIAIKP 565
QY 237 TKGOLVSPVSGKIVVAFPPSGHAFVTRTKAEDGSNVDILMHIGFDVTNLTGTHFNPLKKQG 296
Db 566 SQGEVRAPFNGKVMIFPTKHAIGL-----VSDSGLELLIHIGLDTVKLNGEGFTLHVEEG 621
QY 297 DEVKAGELLCFDDIDAKAAGYEVTTPIVVS 327
Db 622 QEVKQDGLLNFLLDYIRNHAKS DITPIIVT 652
```

RESULT 7  
PCT-US02-03987-12270  
; Sequence 12270, Application PC/TUS0203987  
; GENERAL INFORMATION:  
; APPLICANT: Elittra Pharmaceuticals, Inc.  
; TITLE OF INVENTION: Methods for identifying the Target of a Compound which Inhibits  
; FILE OF INVENTION: Proliferation  
; FILE REFERENCE: ELITRA.028VPC  
; CURRENT APPLICATION NUMBER: PCT/US02/03987  
; CURRENT FILING DATE: 2002-02-02  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09  
; NUMBER OF SEQ ID NOS: 15811  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12270  
; LENGTH: 681  
; TYPE: PRT  
; ORGANISM: Staphylococcus aureus  
PCT-US02-03987-12270

Query Match 17.1%; Score 310.5; DB 1; Length 681;  
Best Local Similarity 24.6%; Pred. No. 2.6e-15;  
Matches 111; Conservative 54; Mismatches 149; Indels 137; Gaps 11;

```
QY 7 VGLLGLVYSPVITGLHOSF-----PPIELELFNOG----- 39
Db 211 VAVFLGFKRLIPFGLHHIFHAPFWFEGSNKNAAGEIHHGDQRIEIQREGAHLTA 270
QY 40 GSFIATASMANIAQAACLAFLAKSEKLGLAGASGVA-----VLGITEPAIFGVNL 95
Db 271 GKFMQGEFFVMFGLPAAALAIYHSAPENKVVVAGLMSAALTSTLTGITEPLEFSFLF 330
QY 96 RLKWPFFIGTGAAGGALIALFNKAVLG-----AAGFLGVVSDA-PDMVME----- 144
Db 331 VAPLLFFI-----HAVLDGLSFLTLVLLDVLHGYTFSGGFDYVLLGLVLPNKQTQWMLVIPVG 387
QY 145 LVCVVVTFE-----AFCAAIAYGLY----- 165
Db 388 LVAVIYVFFRFLVVKYKTPGREDKQSOAVTASATELPYAVLEAMGKGNKIKHLDAC 447
QY 166 LVR-----RNGSIDPD 176
Db 448 ITRLRVEVNDKSKVDVPLKDLGASGVLEGVNNMQAIFGPKSDQIKHEMQOIMNGQVVEN 507
QY 177 ATAAPVPAGTTKAEAPAEFSDNSTIIQAPLTGEATLSSVSDAMPASGKLGSGVAIVP 236
Db 508 PTMEDDKDETIVVVAEDKTSATSELSHIVHAPLTGEVTPLEVPDQVFESEKMMGSGIAIKP 567
QY 237 TKGOLVSPVSGKIVVAFPPSGHAFVTRTKAEDGSNVDILMHIGFDVTNLTGTHFNPLKKQG 296
Db 568 SQGEVRAPFNGKVMIFPTKHAIGL-----VSDSGLELLIHIGLDTVKLNGEGFTLHVEEG 623
QY 297 DEVKAGELLCFDDIDAKAAGYEVTTPIVVS 327
Db 624 QEVKQDGLLNFLLDYIRNHAKS DITPIIVT 654
```

RESULT 8  
US-09-815-242-12270  
; Sequence 12270, Application US/09815242  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; TITLE OF INVENTION: Prokaryotes  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12270  
; LENGTH: 681  
; TYPE: PRT  
; ORGANISM: Staphylococcus aureus  
US-09-815-242-12270

Query Match 17.1%; Score 310.5; DB 6; Length 681;  
Best Local Similarity 24.6%; Pred. No. 2.6e-15;  
Matches 111; Conservative 54; Mismatches 149; Indels 137; Gaps 11;

QY 7 VGGLLFGLVSPVITGLHQSFP-----PPIELFNQ----- 39  
DB 211 VAVFLFGFKIKRLLIPFGLHHIFHAPFWFEGSWKNAAGEIIHQDQRIETIEQIREGAHLTA 270  
QY 40 GSFIFATASMANIAOAGAACLAFLVFLAKSEKLAGAGSVSA-----VLGITPAIFGVNL 95  
DB 271 GKFMGGEPPVMMFGLPAAALAIYHSAPKPNKVVAGLMSAALTSTLTGITEPLEFSEFL 330  
QY 96 RLRPWFIFIGTAAIGGALIALFNKAVAGL---AAGFLGVVSIDA---PDMVNF----- 144  
DB 331 VAPLLFFI---HAVLDGLSFLTYLLDVLHGYTSGGFDIDVLLGVLPNKTQWLVIPIVG 387  
QY 145 LVCVVTFEFI-----AFGAAIAYGLY----- 165  
DB 388 LVYAVIYVVFRLIVKLKYTPGREDKQSOAVTASATELPYAVLEAMGGRKANIKHLDAC 447  
QY 166 LVR-----RNGSIDPD 176  
DB 448 ITRLREVNDSKVDVPGKLDGASGVLEVGNMQAIFGPKSDQIKHEMQQIMNGOVVEN 507  
QY 177 ATAAPVPAGTTKAEAPAEFSNDSTIIQAPLTGAEIALSSVSDAMFASGKLGSVAIYP 236  
DB 508 PTTMEDDKDETIVVNAEDKSATSELSHIVHAPLTGVTPLSEVPDQVFESEKMMGDGIAIKP 567  
QY 237 TKGOLVSPVSGKIVVAPFSGHAFVTRKAEDGSNDVILMHIGFDTVNLNGTHFNPLKKQG 296  
DB 568 SOGEVRAPFNGKVMIPPTKHAIGL---VSDSGLELLIHIGLDTVKLNGEGFTLHVEEG 623  
QY 297 DEVKAGELLCFDDAIIKAAGVEVTTPIVVS 327  
DB 624 QEVKOGDLLINFDLDYIRNHAKSDDITPIIYT 654

RESULT 9  
PCT-US02-851-12270  
; Sequence 12270, Application US/10072851  
; GENERAL INFORMATION:  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Xu, H. Howard  
; APPLICANT: Foulkes, J. Gordon  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Roemer, Terry  
; APPLICANT: Jiang, Bo  
; APPLICANT: Boone, Charles  
; APPLICANT: Bussey, Howard  
; TITLE OF INVENTION: Methods for Identifying the Target of a Compound which Inhibits C  
; FILE REFERENCE: ELITRA.028A  
; CURRENT APPLICATION NUMBER: US/10/072.851  
; CURRENT FILING DATE: 2002-02-08  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09  
; NUMBER OF SEQ ID NOS: 15811  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12270  
; LENGTH: 681  
; TYPE: PRT  
; ORGANISM: Staphylococcus aureus  
US-10-072-851-12270

Query Match 17.1%; Score 310.5; DB 7; Length 681;  
Best Local Similarity 24.6%; Pred. No. 2.6e-15;

Matches 111; Conservative 54; Mismatches 149; Indels 137; Gaps 11;

QY 7 VGGLLFGLVSPVITGLHQSFP-----PPIELFNQ----- 39  
DB 211 VAVFLFGFKIKRLLIPFGLHHIFHAPFWFEGSWKNAAGEIIHQDQRIETIEQIREGAHLTA 270  
QY 40 GSFIFATASMANIAOAGAACLAFLVFLAKSEKLAGAGSVSA-----VLGITPAIFGVNL 95  
DB 271 GKFMGGEPPVMMFGLPAAALAIYHSAPKPNKVVAGLMSAALTSTLTGITEPLEFSEFL 330  
QY 96 RLRPWFIFIGTAAIGGALIALFNKAVAGL---AAGFLGVVSIDA---PDMVNF----- 144  
DB 331 VAPLLFFI---HAVLDGLSFLTYLLDVLHGYTSGGFDIDVLLGVLPNKTQWLVIPIVG 387  
QY 145 LVCVVTFEFI-----AFGAAIAYGLY----- 165  
DB 388 LVYAVIYVVFRLIVKLKYTPGREDKQSOAVTASATELPYAVLEAMGGRKANIKHLDAC 447  
QY 166 LVR-----RNGSIDPD 176  
DB 448 ITRLREVNDSKVDVPGKLDGASGVLEVGNMQAIFGPKSDQIKHEMQQIMNGOVVEN 507  
QY 177 ATAAPVPAGTTKAEAPAEFSNDSTIIQAPLTGAEIALSSVSDAMFASGKLGSVAIYP 236  
DB 508 PTTMEDDKDETIVVNAEDKSATSELSHIVHAPLTGVTPLSEVPDQVFESEKMMGDGIAIKP 567  
QY 237 TKGOLVSPVSGKIVVAPFSGHAFVTRKAEDGSNDVILMHIGFDTVNLNGTHFNPLKKQG 296  
DB 568 SOGEVRAPFNGKVMIPPTKHAIGL---VSDSGLELLIHIGLDTVKLNGEGFTLHVEEG 623  
QY 297 DEVKAGELLCFDDAIIKAAGVEVTTPIVVS 327  
DB 624 QEVKOGDLLINFDLDYIRNHAKSDDITPIIYT 654

RESULT 10  
PCT-US02-03987-10089  
; Sequence 10089, Application PC/TUS0203987  
; GENERAL INFORMATION:  
; APPLICANT: Elittra Pharmaceuticals, Inc.  
; TITLE OF INVENTION: Methods for Identifying the Target of a Compound which Inhibit  
; FILE REFERENCE: ELITRA.028VPC  
; CURRENT APPLICATION NUMBER: PCT/US02/03987  
; CURRENT FILING DATE: 2002-02-02  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09  
; NUMBER OF SEQ ID NOS: 15811  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 10089  
; LENGTH: 648  
; TYPE: PRT  
; ORGANISM: Escherichia coli  
PCT-US02-03987-10089

Query Match 16.2%; Score 295; DB 1; Length 648;  
Best Local Similarity 23.2%; Pred. No. 3.5e-14;  
Matches 114; Conservative 62; Mismatches 162; Indels 154; Gaps 12;

QY 5 GPVGGLLFGLVSPVITGLHQSFPPI-----ELELFNQ-----GSF 42  
DB 170 GALSGIFGFINRLIPLTGLHQLVNTIAWFQIGFTNAAGTVFHGDIAGDGTAGMF 229  
QY 43 IPATASMANIAOAGAACLAFLVFLAKSEK---LKGAGASGVSAVL-GITPPIAF----- 91  
DB 230 MSGFFPIMMFGPLPGAALAMYFAAPKRRPVMVGMLLSVAVTAFLTGTETLEFLMFLAP 289  
QY 92 -----GVNRLRWPFFIGTAAIGGAL----- 114  
DB 290 LLYLLHALITGISLGFVATLLGIHAGFSFAGADYALMNLPAASQNVNMLLMVGVIFFA 349  
QY 115 -----IALFNKA-----VALGAA----- 128

Db 350 IYFVFSLVIRNFKTPGREDKEDEIVTEANSNTEGLTQLATNYIAAVGTDNLKAI 409  
QY 129 -----GFLGVVSIDAPDMVNFELVCVVTFIAFGAAIAYG 163  
Db 410 DACITRLRLTVADARSVNDTMCKRLGASGVVKNKQTIQIVGAKAESIGDAMKKVVARG 469  
QY 164 LYLVRNGSIDPATAAPVPAGTTKAEAEAPAEFSNDSTIIQAPLTGEAIALSSVSDAMF 223  
Db 470 -PVAASAEATP-ATAAPVAKPOAVPNAVSIAE-----LVSPITGDDVVALDOVPDEAF 520  
QY 224 ASKGLSGVAIVPTKQGLVSPVSGKIIVAPPSGHAFVARTKAEADGSDNVDILMHIGFDTVN 283  
Db 521 ASKAVGDGVAVRPTDKIVVSPAAGTIVKIFNTNHAFCLETE---KGAIEIVVHMGIDTVA 576  
QY 284 LMGTHFNPLKOGDEVKAGELCEFDIDAIAKAAGYEVTTPVIVSNYKKTGPNVNTYGLGEI 343  
Db 577 LEGKGFRLVEGAQVSAGQPILEMDLDVNLNANARSMISPVVCSNIDDFSGLIIRKAGGHI 636  
QY 344 EAGANLLNVAKK 355  
Db 637 VAGQTPLYEIKK 648

## RESULT 11

US-09-815-242-10089  
; Sequence 10089, Application US/09815242  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; PRIOR FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 10089  
; LENGTH: 648  
; TYPE: PRT  
; ORGANISM: Escherichia coli  
US-09-815-242-10089

Query Match 16.2%; Score 295; DB 6; Length 648;  
Best Local Similarity 23.2%; Pred. No. 3.5e-14;  
Matches 114; Conservative 62; Mismatches 162; Indels 154; Gaps 12;  
QY 5 GPVGGLLGLVSPVITGLHQSFPPI-----EELFNQG----GSF 42  
Db 170 GALGSGIFGFINRLIPTGLHOVLNTIAWFQIGFTNAAGTVPHGDINRFYAGDGTAGMF 229  
QY 43 IPATASMANIAOGAAACLAFFVFLAKSEK---LKGLAGAGVSAVL-GITEPAIF----- 91

Db 230 MSGFFPIMMFGPLPGAALAMYFAAPKERRPMVGMILLSVAVTAFLTGVTPELEFLMFLAP 289  
QY 92 -----GVNRLRWPFFIGTAAIGGAL----- 114  
Db 290 LLYLLHALLTGISLFTVATLLGIHAGFSFSAGADYALMYNLPAASONVMMLLVMGVIFFA 349  
QY 115 -----IALFNKA-----VALGAA----- 128  
Db 350 IYFVFSLVIRNFKTPGREDKEDEIVTEANSNTEGLTQLATNYIAAVGTDNLKAI 409  
QY 129 -----GFLGVVSIDAPDMVNFELVCVVTFIAFGAAIAYG 163  
Db 410 DACITRLRLTVADARSVNDTMCKRLGASGVVKNKQTIQIVGAKAESIGDAMKKVVARG 469  
QY 164 LYLVRNGSIDPATAAPVPAGTTKAEAEAPAEFSNDSTIIQAPLTGEAIALSSVSDAMF 223  
Db 470 -PVAASAEATP-ATAAPVAKPOAVPNAVSIAE-----LVSPITGDDVVALDOVPDEAF 520  
QY 224 ASKGLSGVAIVPTKQGLVSPVSGKIIVAPPSGHAFVARTKAEADGSDNVDILMHIGFDTVN 283  
Db 521 ASKAVGDGVAVRPTDKIVVSPAAGTIVKIFNTNHAFCLETE---KGAIEIVVHMGIDTVA 576  
QY 284 LMGTHFNPLKOGDEVKAGELCEFDIDAIAKAAGYEVTTPVIVSNYKKTGPNVNTYGLGEI 343  
Db 577 LEGKGFRLVEGAQVSAGQPILEMDLDVNLNANARSMISPVVCSNIDDFSGLIIRKAGGHI 636  
QY 344 EAGANLLNVAKK 355  
Db 637 VAGQTPLYEIKK 648

## RESULT 12

US-10-072-851-10089  
; Sequence 10089, Application US/10072851  
; GENERAL INFORMATION:  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Xu, H. Howard  
; APPLICANT: Foulkes, J. Gordon  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Roemer, Terry  
; APPLICANT: Jiang, Bo  
; APPLICANT: Boone, Charles  
; APPLICANT: Bussey, Howard  
; TITLE OF INVENTION: Methods for Identifying the Target of a Compound which Inhibit  
; FILE REFERENCE: ELITRA.028A  
; CURRENT APPLICATION NUMBER: US/10/072,851  
; CURRENT FILING DATE: 2002-02-08  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09  
; NUMBER OF SEQ ID NOS: 15811  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 10089  
; LENGTH: 648  
; TYPE: PRT  
; ORGANISM: Escherichia coli  
US-10-072-851-10089

Query Match 16.2%; Score 295; DB 7; Length 648;  
Best Local Similarity 23.2%; Pred. No. 3.5e-14;  
Matches 114; Conservative 62; Mismatches 162; Indels 154; Gaps 12;  
QY 5 GPVGGLLGLVSPVITGLHQSFPPI-----EELFNQG----GSF 42  
Db 170 GALGSGIFGFINRLIPTGLHOVLNTIAWFQIGFTNAAGTVPHGDINRFYAGDGTAGMF 229

```

Query Match      15.23; Score 276.5; DB 1; Length 551;
Best Local Similarity 23.09; Pred. No. 6.8e-13;
Matches 107; Conservative 68; Mismatches 130; Indels 161; Gaps 17;

Qy  5  GPVGGLFGLVYSPIVITGLHQSPPI---ELELFNQGGSPF----- 43
      |||  |||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||
Db  77  GFVCAGIVGFENRLLIPTGLHALLNSVFWFVAGINDIGNELLAGOOALDTGKIVGOTGM 136

```

Qy	44	---	FALASHMAIAQQAAGACUAFEEELANSEKALAG	---	ASGVSAYV	---	ILILPAIPVAD	35
Db	137	YQAGFFPVMFGLPAGA	-FAIYQARPEKKVYASLMLAAGFAAFTGVTEPLESF	-M	193			
Qy	96	RLRPFFIGIGTAAIGALIALENIKAVALG	-----AAGFLG-VVSDADP	---	140			
Db	194	FVAMPLIV	-----LHAVTGISLAFEAEMHMTAGFAPSAGFVDFLSLKNPVANH	243				
Qy	141	---MMFLCAVVTFF	-----IAEG	---	AAIAYGLY	165		
Db	244	PMMLVQGLVFAAIYYEGFRFAITKFMLTGPREBGDGETPDVAEGDNKFLARRIYD	303					
Qy	166	---	LVRRNGSIDP	---	---	175		
Db	304	GLGADANVTSIDNCTTRLRLTVKDTGKVDQAKIKATGCPGVKVIDDTNIQIVVGTVEQVF	363					
Qy	176	-----DATAAPV	---PAGTTKAEABAPAEFSNDSTIIQAPUTGERIALSSVS DAM	222				
Db	364	ADEMQRLYNHQAPATPVKETPTQSPVVPVEKAPV	-STKETELYSVANGKVIPISEYPDVV	421				
Qy	223	FASRKLGSVAIVPTKQLVSPVSGKIVVAFPSGHAFAVETKAEKDSNVDDLWHIGHGFDTV	282					
Db	422	FSAKNMGDGEAVVPTDGVSTPVAGKITSIFPTKHALGIQT	-----DSGIEVLULHMGIDIV	477				
Qy	283	NLNGTHENPLKQGDVEYKAGSELICEFDIDAIIKAAGYEVITTPVYSN	328					
Db	478	FLOGGPETLHVEFGVGVKGDKTATIDLAALFOAGKKSDILVVFTN	523					

RESULT 15  
US-09-815-242-10500  
; Sequence 10500, Application US/09815242  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Karl L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T

APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; PRIOR FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 10500  
; LENGTH: 551  
; TYPE: PRT  
; ORGANISM: Enterococcus faecalis  
US-09-815-242-10500

Query Match 15.2%; Score 276.5; DB 6; Length 551;  
Best Local Similarity 23.0%; Pred. No. 6.8e-13;  
Matches 107; Conservative 68; Mismatches 130; Indels 161; Gaps 17;  
  
QY 5 GPGVGLLFLGLVSPVITGLHQSFPI---EELFNQGSFI----- 43  
Db 77 GFVGAGLYGFNNLLIPTGLHNLNSVFWDFVAGINDIGNFLAGQQALDTGKAIVGQTGM 136  
  
QY 44 ---FATASMANIAQGAACLAFFLAKSEKLGKLAG---ASGVSAVL-GITEPAIFGVNL 95  
Db 137 YQAGFFPVMFGLPAGA--FAIQCARPEKKVKTASIMLAAGFAAFFGVTEPLEFSF-M 193  
  
QY 96 RLKWPFFIGTAAIGGALIALNIKVALG-----AAGFLG-VVSDAPD--- 140  
Db 194 FVAMPVLYV-----LHAVFTGISLAPAEPMHWNTAGFAFSAGFVDFFLSLKNPVANH 243  
  
QY 141 ---MYMFLVCVVTFE- - - - -IAFG---AAIAYGLY- 165  
Db 244 PMMLVQGLVFAAIYFGFRFAITKFNLMTPGREEGDGETPDVAGDNKPFASLARIYD 303  
  
QY 166 -----LVRNGSIDP----- 175  
Db 304 GLGADANVTSIDNCTTRLRLTVKDTGKVQDAKIKATGVPGVKVIDDTNIQIVIGTEVQFV 363  
  
QY 176 -----DATAAPV---PAGTTKAEAPAPAEFNSDSTIIQAPLTGEAIALSSVSAM 222  
Db 364 ADEMQRLYNHQADPATPVKETPVSQPVVERKAPV--STKETELYSVANGKVIPISEVPDDV 421  
  
QY 223 FASSKLGSGVAIVPTKGOLVSPVSGKIVVAFPSGSHAFVARTKAEDGSNDVILMHIGFDTV 282  
Db 422 FSAKMGDGFVAIVPTDGVSTPVAGKITSTIFPTKHALGIQT-----DSGIEVLHLMGLDTV 477  
  
QY 283 NLNGTHFNPLKQGVDEKAGELCEFDIDAIAAGYEVTTPPIVVSN 328  
Db 478 ELQGGPFTLHVEESGVVQKQDKIATIDLAALEQAGKKSDLIVVFTN 523

Search completed: March 21, 2002, 16:27:37  
Job time: 338 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 21, 2002, 16:23:16 ; Search time 21.58 Seconds  
(without alignments)  
377.488 Million cell updates/sec

Title: US-09-604-231-4  
Perfect score: 1821  
Sequence: 1 YDFGPPVGLLFLGLVSPV.....IERGANLLNVAKREAVPATP 362

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents-AA:\*  
1: /cgn2\_6/ptodata/2/1aa/5A-COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B-COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6A-COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6B-COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/PCTUS-COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/backfiles1.pep:\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	113.5	6.2	1176	1	US-07-828-788A-12
2	113.5	6.2	1176	1	Sequence 12, Appl
3	113.5	6.2	1176	3	Sequence 2, Appl1
4	113.5	6.2	1176	5	Sequence 2, Appl1
5	113.5	6.2	1179	1	Sequence 12, Appl
6	106.5	5.8	1129	6	Sequence 1, Appl1
7	106.5	5.8	1179	6	Patent No. 5164180
8	99.5	5.5	522	4	Patent No. 5188960
9	99.5	5.5	654	4	Sequence 3, Appl1
10	98	5.4	1184	6	Sequence 35, Appl
11	98	5.4	1188	6	Patent No. 5254799
12	97.5	5.4	2005	3	Patent No. 5254799
13	96.5	5.3	593	6	Sequence 7, Appl1
14	96.5	5.3	648	3	Patent No. 5523211
15	96.5	5.3	934	1	Sequence 13, Appl
16	96.5	5.3	934	1	Sequence 7, Appl1
17	96.5	5.3	1011	3	Sequence 2, Appl1
18	96.5	5.3	1176	1	Sequence 6, Appl1
19	96.5	5.3	1176	1	Sequence 10, Appl
20	96.5	5.3	1176	1	Sequence 10, Appl
21	96.5	5.3	1176	5	Sequence 6, Appl1
22	96.5	5.3	1176	5	Sequence 10, Appl
23	96.5	5.3	1984	3	Sequence 10, Appl
24	96.5	5.3	1989	3	Sequence 12, Appl
25	95	5.2	680	2	Sequence 12, Appl
26	94	5.2	1835	3	Sequence 2, Appl1
27	94	5.2	2259	4	Sequence 15, Appl
					Sequence 70, Appl

28 94 5.2 2439 3 US-09-335-409-7 Sequence 7, Appl1  
29 92.5 5.1 614 2 US-08-622-740-8 Sequence 8, Appl1  
30 92.5 5.1 614 3 US-08-440-689-8 Sequence 8, Appl1  
31 92.5 5.1 615 3 US-09-042-426-10 Sequence 10, Appl  
32 92.5 5.1 615 4 US-09-291-238-10 Sequence 10, Appl  
33 92.5 5.1 615 4 US-09-330-760-10 Sequence 10, Appl  
34 92.5 5.1 615 4 US-09-328-473-10 Sequence 10, Appl  
35 92.5 5.1 615 4 US-09-330-737-10 Sequence 10, Appl  
36 92.5 5.1 695 1 US-07-671-817A-5 Sequence 5, Appl1  
37 92.5 5.1 969 1 US-07-671-817A-6 Sequence 6, Appl1  
38 92.5 5.1 1155 1 US-08-349-867-19 Sequence 13, Appl  
39 92.5 5.1 1155 1 US-08-349-867-33 Sequence 33, Appl  
40 92.5 5.1 1155 1 US-08-239-476-19 Sequence 19, Appl  
41 92.5 5.1 1155 1 US-08-446-486-2 Sequence 2, Appl1  
42 92.5 5.1 1155 1 US-07-951-715A-9 Sequence 9, Appl1  
43 92.5 5.1 1155 1 US-08-463-308-2 Sequence 2, Appl1  
44 92.5 5.1 1155 2 US-08-598-305A-19 Sequence 19, Appl  
45 92.5 5.1 1155 2 US-08-598-305A-33 Sequence 33, Appl

#### ALIGNMENTS

RESULT 1  
US-07-828-788A-12  
: Sequence 12, Application US/07828788A  
: Patent No. 5273746  
: GENERAL INFORMATION:  
: APPLICANT: PAYNE, JEWEL M.  
: APPLICANT: HICKLE, LESLIE A.  
: TITLE OF INVENTION: NOVEL BACILLUS THURINGIENSIS ISOLATES  
: TITLE OF INVENTION: ACTIVE AGAINST PHTHIRAPTERA PESTS  
: NUMBER OF SEQUENCES: 16  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: DAVID R. SALIWANCHIK  
: STREET: 2421 N.W. 41st STREET, SUITE A-1  
: CITY: GAINESVILLE  
: STATE: FL  
: COUNTRY: USA  
: ZIP: 32606  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: Patentin Release #1.0, Version #1.25  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/07/828,788A  
: FILING DATE: 19920129  
: CLASSIFICATION: 435  
: ATTORNEY/AGENT INFORMATION:  
: NAME: SALIWANCHIK, DAVID R.  
: REGISTRATION NUMBER: 31,794  
: REFERENCE/DOCKET NUMBER: MAY5  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: 904-375-8100  
: TELEFAX: 904-372-5800  
: INFORMATION FOR SEQ ID NO: 12:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 1176 amino acids  
: TYPE: AMINO ACID  
: STRANDEDNESS: single  
: TOPOLOGY: linear  
: MOLECULE TYPE: protein  
: HYPOTHETICAL: YES  
: ANTI-SENSE: NO  
: ORIGINAL SOURCE:  
: ORGANISM: BACILLUS THURINGIENSIS  
: STRAIN: AIZAWAI  
: INDIVIDUAL ISOLATE: PS811  
: IMMEDIATE SOURCE:  
: LIBRARY: LAMBDA GEM (TM) - 11 LIBRARY OF AUGUST SICK  
: CLONE: 811A2  
: US-07-828-788A-12

```

; TELEFAX: (904)372-5800
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1176 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: BACILLUS THURINGIENSIS
; STRAIN: AIZAWAI
; INDIVIDUAL ISOLATE: PS81I
; IMMEDIATE SOURCE:
; LIBRARY: LAMBDA GEM (TM) - 11 LIBRARY OF AUGUST SICK
; CLONE: 811A2
; US-08-356-034-2

Query Match      6.2%; Score 113.5; DB 1; Length 1176;
Best Local Similarity 22.4%; Pred. No. 0.0076;
Matches 56; Conservative 43; Mismatches 96; Indels 55; Gaps 13;

QY 143 MFLCAVVTFFIAFGAAIAYGLYLVRRNGSID-----PDATAAPVPAG-----TTK 188
Db 377 LFVLDTGTEFSFASLTADLPSTIY--RQGTVDLSLDVIPQDINSVPARAGFSHRLSHVTML 434
QY 189 AEA-----EAP-----AEFSN---DSTTIQAPLTGEAIALSSVSDAMFASGKLG 230
Db 435 SQAAGAVYTLRAPTFWSHRHSAEFSNLPSSQITQIPLT-KSINLGSCTSVKPGPGFTGG 493
QY 231 GVAIVPTKGOLVSPVSGKIVVAPSGHAFVTRKAEDGSNVDILMHIGFDFTVNLNGTHEN 290
Db 494 DILRITSPGQISTL---RVITAPLSQRYVRIRY--ASTTNLQFHTSIDGRPINQGNFS 548
QY 291 PLKQSGDEVKAGELLCEFDIDAIAKAAGYEVYTPVWSNYKTKGTPVNTYGLGEIEAGANLL 350
Db 549 ATMSSGGNLSQSG-----SPRTAGF--TTPE---NFSNGSSIFTLSAHVFNNG-NEV 593
QY 351 NVAKKEAVPA 360
Db 594 YIERIEFVPA 603

RESULT 3
US-08-933-891-2
; Sequence 2, Application US/08933891
; Patent No. 6096708
; GENERAL INFORMATION:
; APPLICANT: Payne, Jewel M.
; APPLICANT: Sick, August J.
; TITLE OF INVENTION: No. 6096708el Bacillus thuringiensis Isolate
; TITLE OF INVENTION: Active Against Lepidopteran Pests, and Genes Encoding
; TITLE OF INVENTION: No. 6096708el Lepidopteran-Active Toxins
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: US
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/933,891
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/371,955
; FILING DATE: 27-JUN-89
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Saliwanchik, Roman
; REGISTRATION NUMBER: 21,023
; REFERENCE/DOCKET NUMBER: MA43.C1.D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (904)375-8100
```





## GENERAL INFORMATION:

APPLICANT: Payne, Jewel M  
APPLICANT: Sick, August J  
TITLE OF INVENTION: No. 5407825el Bacillus thuringiensis isolates  
TITLE OF INVENTION: active against Lepidopteran Pests and Genes Encoding No. 54078  
TITLE OF INVENTION: Lepidopteran-active Toxins  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: DAVID R. SALIWANCHIK  
STREET: 2421 N.W. 41st STREET, SUITE A-1  
CITY: GAINESVILLE  
STATE: FL  
COUNTRY: USA  
ZIP: 32606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/040,751  
FILING DATE: 19930329  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: SALIWANCHIK, DAVID R.  
REFERENCE/DOCKET NUMBER: MA39.C1.D3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 904-375-8100  
TELEFAX: 904-372-5800  
TELEX:  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1179 amino acids  
TYPE: AMINO ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: BACILLUS THURINGIENSIS  
STRAIN: AIZAWAI  
INDIVIDUAL ISOLATE: PS81I  
IMMEDIATE SOURCE:  
LIBRARY: LAMBDA GEM (TM) - 11 LIBRARY OF AUGUST SICK  
CLONE: 81LR1  
US-08-040-751-1

Query Match 6.2%; Score 113.5; DB 1; Length 1179;  
Best Local Similarity 22.4%; Pred. No. 0.0076;  
Matches 56; Conservative 43; Mismatches 96; Indels 55; Gaps 13;

QY 143 MFLVCVVVTFIAFGAAIAYGLYLVRNGSID-----PDATAAPVPAG-----TTK 188  
Db 380 LFVLDGTEFSFASLTADLPSTIY--RQGTVDLSLDVIPPDNSVPARAGFSHRLSHVTML 437  
QY 189 AEA-----EAP-----AEFSN---DSTIIQAPLTGEATLSSVSDAMFASGKLG 230  
Db 438 SQAAGAVYTLRAPTFSWRHSAEFSLIPSSQITQIPLT-KSINLGSSTVYKPGFTGG 496  
QY 231 GVAIVPTKGQVSPVSGKIVVAPPSGHAFVTRTKAEDGSNVDILMHGFTVNLNGTHFN 290  
Db 497 DILRITSPGQISTL---RVITAPLSQRYVRIRY--ASTTNLQFHTSIDGRPINQGNFS 551  
QY 291 PLKKQDEVKAGELLCEFDIDAIAKAGYEVVTPIVVSNYKTKGPNVYGLGEIAGANLL 350  
Db 552 ATMSSGGNLSQSG-----SFRTAGF--TTFP---NFSNGSSIETLSAHVFNESG-NEV 596  
QY 351 NVAKKEAVPA 360  
Db 597 YIERIEFVPA 606

RESULT 6  
5164180-6  
Patent No. 5164180  
APPLICANT: Payne, Jewel; Sick, August J.  
TITLE OF INVENTION: BACILLUS THURINGIENSIS ISOLATES ACTIVE  
AGAINST LEPIDOPTERAN PESTS  
NUMBER OF SEQUENCES: 6  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/451,389  
FILING DATE: 14-DEC-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 353,860  
FILING DATE: 18-MAY-1989  
SEQ ID NO: 6:  
LENGTH: 1129  
5164180-6

Query Match 5.8%; Score 106.5; DB 6; Length 1129;  
Best Local Similarity 22.4%; Pred. No. 0.039;  
Matches 56; Conservative 42; Mismatches 97; Indels 55; Gaps 13;

QY 143 MFLVCVVVTFIAFGAAIAYGLYLVRNGSID-----PDATAAPVPAG-----TTK 188  
Db 380 LFVLDGTEFSFASLTADLPSTIY--RQGTVDLSLDVIPPDNSVPARAGFSHRLSHVTML 437  
QY 189 AEA-----EAP-----AEFSN---DSTIIQAPLTGEATLSSVSDAMFASGKLG 230  
Db 438 SQAAGAVYTLRAPTFSWRHSAEFSLIPSSQITQIPLT-KSINLGSSTVYKPGFTGG 496  
QY 231 GVAIVPTKGQVSPVSGKIVVAPPSGHAFVTRTKAEDGSNVDILMHGFTVNLNGTHFN 290  
Db 497 DILRITSPGQISTL---RVITAPLSQRYVRIRY--ASTTNLQFHTSIDGRPINQGNFS 551  
QY 291 PLKKQDEVKAGELLCEFDIDAIAKAGYEVVTPIVVSNYKTKGPNVYGLGEIAGANLL 350  
Db 552 ATMSSGGNLSQSG-----SFRTAGF--TTFP---NFSNGSSIETLSAHVFNESG-NEV 596  
QY 351 NVAKKEAVPA 360  
Db 597 YIERIEFVPA 606

RESULT 7  
5188960-2  
Patent No. 5188960  
APPLICANT: PAYNE, JEWEL; SICK, AUGUST J.  
TITLE OF INVENTION: BACILLUS THURINGIENSIS ISOLATE ACTIVE  
AGAINST LEPIDOPTERAN PESTS, AND GENES ENCODING NOVEL  
LEPIDOPTERAN-ACTIVE TOXINS  
NUMBER OF SEQUENCES: 8  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/451,261  
FILING DATE: 14-DEC-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 371,955  
FILING DATE: 27-JUN-1989  
SEQ ID NO: 2:  
LENGTH: 1179  
5188960-2

Query Match 5.8%; Score 106.5; DB 6; Length 1179;  
Best Local Similarity 22.4%; Pred. No. 0.042;  
Matches 56; Conservative 42; Mismatches 97; Indels 55; Gaps 13;

QY 143 MFLVCVVVTFIAFGAAIAYGLYLVRNGSID-----PDATAAPVPAG-----TTK 188  
Db 380 LFVLDGTEFSFASLTADLPSTIY--RQGTVDLSLDVIPPDNSVPARAGFSHRLSHVTML 437  
QY 189 AEA-----EAP-----AEFSN---DSTIIQAPLTGEATLSSVSDAMFASGKLG 230

Db 438 SOAGAVYTLRAPTSWRHRSAEFNSLIPSSQITQIPLT-KSINLGSCTSVVKGPGFTGG 496  
QY 231 GVAIVPTKQLVSPVSGKIIVAFPSGSHAFVATKAEKSDNDILMHIGFDIVNLGTHEN 290  
Db 497 DILRTPSQISTL---RVITAPLSQRYRIRY--ASTNLQFHTSIDGRPINQGNFS 551  
QY 291 PLKKGDEVKAGELLCEFDIDAIAKAAGYEVTTPIVVSNKKTGPNVTYGLGIEAGANLL 350  
Db 552 ATMSSGGNLQSG-----SFRTAGF--TTPF---NFSNGSSIFTLSAHVFNSSG-NEV 596  
QY 351 NVAKKEAVPA 360  
Db 597 YIDRIEFVPA 606

RESULT 8  
US-08-894-818B-3  
; Sequence 3, Application US/08894818B  
; Patent No. 6261822  
; GENERAL INFORMATION:  
; APPLICANT: TAKAKURA, Hikaru  
; APPLICANT: MORISHITA, Mio  
; APPLICANT: YAMAMOTO, Katsuhiko  
; APPLICANT: MITTA, Masanori  
; APPLICANT: ASADA, Kiyozo  
; APPLICANT: TSUNASAWA, Susumu  
; APPLICANT: KATO, Ikunoshin  
; TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES  
; NUMBER OF SEQUENCES: 42  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Browdy and Neimark  
; STREET: 419 Seventh Street N.W., Ste. 300  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: United States of America  
; ZIP: 20004  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/894.818B  
; FILING DATE: 20-MAY-1998  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/JP96/03253  
; FILING DATE: 07-NOV-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 323285/1995  
; FILING DATE: 12-DEC-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Browdy, Roger L.  
; REGISTRATION NUMBER: 25,618  
; REFERENCE/DOCKET NUMBER: TAKAKURA-1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 628-5197  
; TELEFAX: (202) 737-3528  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 522 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; FEATURE:  
; OTHER INFORMATION: /note= Xaa at position 428 is Gly or Val.  
US-08-894-818B-3

Query Match 5.5%; Score 99.5; DB 4; Length 522;  
Best Local Similarity 20.6%; Pred. No. 0.063;  
Matches 81; Conservative 63; Mismatches 114; Indels 135; Gaps 21;

QY 1 YDFGPGVGLLFLVYSPVITGLHQSPPPIELEL-----FNQGSFIFAT-----ASMAN 51  
Db 19 WNLGVDGSGITIG-----IIDTGIDASHPDLOGKVIWVDFVNGRSYPTDDHGHGTHVAS 73  
QY 52 TAQGAACLAVFELAKSEKLKGLA-----GASGVSAVLGITEPAI-----FG 92  
Db 74 TAAGTGA-----ASNGKYKGMAPCAKLAGIKVLGADGSGSISTIIKGVEMAVDNKDKYG 127  
QY 93 ---VNLRLWPFFIGI-----GTAAGGALIALFNKAVKALGAAGFLG-----132  
Db 128 IKVINLSL-----GSSQSSDGTALSOAVNAADAGLVVVVAAGNSGPNKYITIGSPAAA 181  
QY 133 ---VVSIDAPDMVFLVCAVVTFFIAFG-----AAIAYGLYLV--RRNGS-----ID 174  
Db 182 SKVITVGAVDKY-----DVITFSRSRGPTADGRLKPEVVPVAPGNWIIAARASGTSMGQPIN 236  
QY 175 PDATAAP-----VP--AGTTKRAEAPAEFSNDSTIIQAPLTGEAIALSSVSDAMFASGK 227  
Db 237 DYYTAAPGTSMAHPVAGIAALLLQAHPSWTPDKVKYKTALLETADIVKPDDEIADIAYGAGR 296  
QY 228 LGSVVAI-----VPTKGQLVSPVSGKIVVAFPPSGHAFVATKAEKSDNDIL 274  
Db 297 VNAYKAINVDNYAKLVFTGYVANKGS---QTHQFVI---SGASEFTATLYWDNANSOLD 349  
QY 275 MHI-----GEDTVNLNGTHENP 291  
Db 350 LYLDPNGNQVDYSYATYGYFEKVG-----YYPN 378

RESULT 9  
US-08-894-818B-35  
; Sequence 35, Application US/08894818B  
; Patent No. 6261822  
; GENERAL INFORMATION:  
; APPLICANT: TAKAKURA, Hikaru  
; APPLICANT: MORISHITA, Mio  
; APPLICANT: YAMAMOTO, Katsuhiko  
; APPLICANT: MITTA, Masanori  
; APPLICANT: ASADA, Kiyozo  
; APPLICANT: TSUNASAWA, Susumu  
; APPLICANT: KATO, Ikunoshin  
; TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES  
; NUMBER OF SEQUENCES: 42  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Browdy and Neimark  
; STREET: 419 Seventh Street N.W., Ste. 300  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: United States of America  
; ZIP: 20004  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/894.818B  
; FILING DATE: 20-MAY-1998  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/JP96/03253  
; FILING DATE: 07-NOV-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 323285/1995  
; FILING DATE: 12-DEC-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Browdy, Roger L.  
; REGISTRATION NUMBER: 25,618  
; REFERENCE/DOCKET NUMBER: TAKAKURA-1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 628-5197  
; TELEFAX: (202) 737-3528



Db 543 FHTSDGRPNQGNFSAWSSGSLQSG-----SFRVGF--TTPF---NFSNGSS 588

QY 335 VNTYGLGEIEAGANLLNVAKEAVPA 360

Db 589 VFTLSARHNSG-NEVYIDRIEFVPA 613

## RESULT 12

US-08-836-325-7

; Sequence 7, Application US/08836325

; Patent No. 6110672

; GENERAL INFORMATION:

; APPLICANT: Mandel, Gail

; APPLICANT: Haleboua, Simon

; APPLICANT: Borden, Laurence A.

; TITLE OF INVENTION: Peripheral Nervous System Specific

; TITLE OF INVENTION: Sodium Channels, DNA Encoding Therefor, Crystallization,

; TITLE OF INVENTION: X-ray Diffraction, Computer Molecular Modeling, Rational

; TITLE OF INVENTION: Drug Design, Drug Screening, and Methods of Making and Using

; NUMBER OF SEQUENCES: 19

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C

; STREET: 1100 New York Ave., N. W., Suite 600

; CITY: Washington

; STATE: DC

; COUNTRY: USA

; ZIP: 20005-3934

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/836,325

; FILING DATE: 2-MAY-1997

; CLASSIFICATION: 514

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/US95/14251

; FILING DATE: 02-NOV-1995

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/482,401

; FILING DATE: 07-JUN-1995

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/334,029

; FILING DATE: 02-NOV-1994

; ATTORNEY/AGENT INFORMATION:

; NAME: Ludwig, Steven R.

; REGISTRATION NUMBER: 36,203

; REFERENCE/DOCKET NUMBER: 0917.0240002

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 202-371-2600

; TELEFAX: 202-371-2540

; INFORMATION FOR SEQ ID NO: 7:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2005 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

US-08-836-325-7

## Query Match

Best Local Similarity 5.4%; Score 97.5; DB 3; Length 2005;

Matches 64; Conservative 42; Mismatches 111; Indels 71; Gaps 12;

QY 70 LKLAGAGCSAV---LGITEPAIFGVNLRWPFPI---GIGTAA-----109

Db 1637 LRLGAKGIRTLFLALMSLPAFGNIGLLFLFYAIFGMSNFAYVKREVIGDDMFN 1696

QY 110 ---IGGALIALFNKAVKALGAAGFLG-VVSDAPD-----WVM 143

Db 1697 FETFGNSMCLFQITISA-GWDGLLAPILNSGPPDCDPEKDHGSSVKGDCGNPSVGIF 1755

QY 144 FLVCVVVTFIAFGAIAAGLYLVRNRGSDPDATAAPVAGTTKAEAPAEFSDSTI 203

Db 1756 FVSYIIISFLVVNMVIA---VILENFSVATEESAPELSEDDFEMFYEWKEKFDPA 1810

QY 204 IQAPLTGEAIALSSVSDAMFASGKLGSGVAIVPTKGQVLS---PVSGKIVVAFPSGHAF 259

Db 1811 -----QFIECKLSD--FAALDPPLIAKPNKVOLIAMDLPMVSGDRIHCLDLILFAP 1861

QY 260 AVRTKAEDG--SNVDILMHIGFDVNLNGTHFNP-----LKKQGEVKA 301

Db 1862 TKRVLGESGEMDALRIQMEEREMASNPMSKVSVEPIITTLKRRQEEVSA 1909

## RESULT 13

5523211-1

; Patent No. 5523211

; APPLICANT: PUSZTAI-CAREY, MARIANNE; CAREY, PAUL R.; LESSARD,

; TIMOTHY; YAGUCHI, MAKOTO

; TITLE OF INVENTION: IDENTIFICATION, QUANTITATION AND

; PURIFICATION OF INSECTICIDAL PROTEINS FROM BACILLUS THURINGIENSIS

; NUMBER OF SEQUENCES: 3

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/277,076

; FILING DATE: 19-JUL-1994

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 102,491

; FILING DATE: 05-AUG-1993

; APPLICATION NUMBER: 836,967

; FILING DATE: 19-FEB-1992

; APPLICATION NUMBER: 493,453

; FILING DATE: 14-MAR-1990

; SEQ ID NO: 1:

; LENGTH: 593

5523211-1

Query Match 5.3%; Score 96.5; DB 6; Length 593;  
Best Local Similarity 22.4%; Pred. No. 0.16; Matches 53; Conservative 33; Mismatches 78; Indels 73; Gaps 13;

QY 166 LVRRNGSID-----PDATAAPVPAG-----TTKAEA-----EAP-----AE 196

Db 373 IYRQGTVDSDLVIPQDNSVPPRAGFHRLSHVTLMSQAAGAVYTLRAFTSWOHSRAE 432

QY 197 FSN---DSTIIQAPLTGEAIALSSVSDAMFASGKLGSGVAIVP-----TKQLVSPVSG-- 247

Db 433 FNNIIPSSQITQIPLT-----KSTNLGSGTSVVKPGFTGGDILRRTSPQG 478

QY 248 ----KIVVAFPSGHAFVTRKAEDGSNVDILMHIGFDVNLNGTHFNP LKKQGEVKA 303

Db 479 ISTLRVNITAPLSQRYVRIRY--ASTTNLQFHTSIDGRPINOGNFSATMSSGSLQSG- 535

QY 304 LLCEFDIDAKAAGYEVTTPIVVSNYKTKGTPVNTYGLGEIEAGANLLNVAKEAVPA 360

Db 536 -----SFRVGF--TTPF---NFSNGSSVFTLSAHVNSG-NEVYIDRIEFVPA 578

## RESULT 14

US-08-810-720-13

; Sequence 13, Application US/08810720

; Patent No. 6037527

; GENERAL INFORMATION:

; APPLICANT: Barton, Kenneth A.

; APPLICANT: Umbeck, Paul F.

; TITLE OF INVENTION: INSECTICIDAL COTTON PLANTS (AS AMENDED)

; NUMBER OF SEQUENCES: 13

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Nicholas J. Seay

; STREET: One South Pinckney Street

; CITY: Madison

; STATE: WI

COUNTRY: US  
ZIP: 53701-2113  
COMPUTER: IBM PC compatible  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/810,720  
FILING DATE: 03-MAR-1997  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: Seay, Nicholas J.  
REGISTRATION NUMBER: 27,386  
REFERENCE/DOCKET NUMBER: 670513.90163  
TELEPHONE: 608/251-5000  
TELEFAX: 608/251-9166  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 648 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-810-720-13

Query Match 5.3%; Score 96.5; DB 3; Length 648;  
Best Local Similarity 22.4%; Pred. No. 0.18;  
Matches 53; Conservative 33; Mismatches 78; Indels 73; Gaps 13;  
QY 166 LVRRNGSID-----PDATAAPVPAG-----TTKAEA-----EAP-----AE 196  
Db 401 IYRQGTVDSDLVIPPDQNSVPPRAGFSLHSHVMTLSQAAGAVYTLRAPTFWSQHRSAE 460  
QY 197 FSN---DSTIIQAPLTGEATLSSVDAMFASGLKLGSGVAIVP-----TKQLVSPVSG-- 247  
Db 461 FNNIIPSSQITQIPLT-----KSTNLGSGTSVVKGPFGTGGDILRTSPGQ 506  
QY 248 ----KIVVAPPSGHAFVARTKAEDGSNVDILMHIGFTVNLNGTHFNPLKKQGEVKAGE 303  
Db 507 ISTRVNITAPLSQRYVRIRY--ASTNLQFHTSIDGRPINQGNFSATMSSGNSNLQSG- 563  
QY 304 LLCEFDIDAIAKAGYEVTTPIVVSNYKKTGPVNTYGLGEIEAGANLLNVAKKEAVPA 360  
Db 564 -----SPTVGF--TTPF---NFSNGSSVFTLSAHVFNNSG-NEVYIDRIEFVPA 606

RESULT 15  
US-08-446-486-7  
Sequence 7, Application US/08446486  
Patent No. 5345565  
GENERAL INFORMATION:  
APPLICANT: De Greve, Henri Marcel J  
APPLICANT: Salgado, Maria Benita Leonor F  
APPLICANT: Van Montagu, Marc Charles E  
APPLICANT: Vaeck, Mark A  
APPLICANT: Zabeau, Marcus Florent O  
APPLICANT: Leemans, Jan Jozef A  
APPLICANT: Hofte, Hermanus Franciscus P  
TITLE OF INVENTION: TRANSFORMATION VECTORS ALLOWING  
TITLE OF INVENTION: EXPRESSION OF FOREIGN POLYPEPTIDE ENDOTOXINS FROM BACILLUS  
TITLE OF INVENTION: THURINGIENSIS IN PLANTS  
NUMBER OF SEQUENCES: 44  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: George Mason Bldg., Washington & Prince Sts.  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/446,486  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA: US 08/133,965  
FILING DATE: 08-OCT-1993  
APPLICATION NUMBER: US 08/014,148  
FILING DATE: 05-FEB-1993  
PRIOR APPLICATION DATA: US 07/555,828  
FILING DATE: 23-JUL-1990  
PRIOR APPLICATION DATA: US 06/821,582  
FILING DATE: 22-JAN-1986  
PRIOR APPLICATION DATA: US 06/692,759  
FILING DATE: 18-JAN-1985  
ATTORNEY/AGENT INFORMATION:  
NAME: Crane-Feury, Sharon E  
REGISTRATION NUMBER: 36,113  
REFERENCE/DOCKET NUMBER: 010830-047  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 934 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
FEATURE:  
NAME/KEY: Protein  
LOCATION: 1..934  
OTHER INFORMATION: /note= "deduced amino acid sequence  
OTHER INFORMATION: of B.t. sotto (Shibano et al., Gene 34, p. 243,  
OTHER INFORMATION: 1985)"  
US-08-446-486-7

Query Match 5.3%; Score 96.5; DB 1; Length 934;  
Best Local Similarity 22.4%; Pred. No. 0.32;  
Matches 53; Conservative 33; Mismatches 78; Indels 73; Gaps 13;  
QY 166 LVRRNGSID-----PDATAAPVPAG-----TTKAEA-----EAP-----AE 196  
Db 401 IYRQGTVDSDLVIPPDQNSVPPRAGFSLHSHVMTLSQAAGAVYTLRAPTFWSQHRSAE 460  
QY 197 FSN---DSTIIQAPLTGEATLSSVDAMFASGLKLGSGVAIVP-----TKQLVSPVSG-- 247  
Db 461 FNNIIPSSQITQIPLT-----KSTNLGSGTSVVKGPFGTGGDILRTSPGQ 506  
QY 248 ----KIVVAPPSGHAFVARTKAEDGSNVDILMHIGFTVNLNGTHFNPLKKQGEVKAGE 303  
Db 507 ISTRVNITAPLSQRYVRIRY--ASTNLQFHTSIDGRPINQGNFSATMSSGNSNLQSG- 563  
QY 304 LLCEFDIDAIAKAGYEVTTPIVVSNYKKTGPVNTYGLGEIEAGANLLNVAKKEAVPA 360  
Db 564 -----SPTVGF--TTPF---NFSNGSSVFTLSAHVFNNSG-NEVYIDRIEFVPA 606

Search completed: March 21, 2002, 16:23:18  
Job time: 149 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 21, 2002, 16:22:45 ; Search time 42.58 Seconds

(without alignments)  
629.745 Million cell updates/sec

Title: US-09-604-231-4

Perfect score: 1821

Sequence: 1 YDFGPGVGLLGLVYSPV.....IEAGANLLNVAKEAVPATP 362

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_1101.\*

```

1: /SID88/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SID88/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SID88/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SID88/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SID88/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SID88/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SID88/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SID88/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SID88/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SID88/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SID88/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SID88/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SID88/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SID88/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SID88/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SID88/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SID88/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SID88/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SID88/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SID88/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SID88/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SID88/gcgdata/geneseq/geneseq/AA2001.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1821	100.0	362	22	AAB66708 C. glutamicum phosph
2	1821	100.0	468	22	AAB66707 C. glutamicum phosph
3	1821	100.0	661	22	AAG92650 C. glutamicum prote
4	1806	99.2	661	22	AAB69080 Brevibacterium lac
5	390.5	21.4	683	22	AAG93207 C. glutamicum prote
6	390.5	21.4	683	22	AAB66721 C. glutamicum phosph
7	295	16.2	648	22	AAG98282 Escherichia coli p
8	234	12.9	135	22	AAB66723 C. glutamicum phosph
9	233.5	12.8	167	20	AA120012 B. burgdorferi ant
10	221.5	12.2	583	21	AA153620 Amino acid sequenc
11	166.5	9.1	135	20	AA120013 B. burgdorferi ant

12	161.5	8.9	126	18	AAW28051	Amino acid sequenc
13	154	8.5	108	21	AAB40791	Human OREF ORF555
14	150.5	8.3	449	21	AA181556	Streptococcus pneu
15	118.5	6.5	805	22	AAG90449	C. glutamicum prote
16	113.5	6.2	1176	14	AA139755	Delta endotoxin.
17	112	6.2	1951	22	AA199678	Human adult form o
18	112	6.2	1951	22	AA199679	Human neonatal for
19	108.5	6.0	806	20	AA121797	Alteromonas lipase
20	108.5	6.0	809	20	AA121796	Alteromonas lipase
21	106.5	5.8	1179	11	AA182558	B. thuringiensis to
22	106.5	5.8	1179	13	AA125826	Novel toxin expres
23	105	5.8	404	22	AA191953	C. glutamicum prote
24	105	5.8	1291	19	AA159912	Amino acid sequenc
25	102.5	5.6	474	19	AA158862	T. halophilus xyl
26	102	5.6	655	22	AA190809	C. glutamicum prote
27	101.5	5.6	405	22	AA15136	Peptide #1570 enco
28	101.5	5.6	405	22	AA127591	Peptide #1628 enco
29	101.5	5.6	405	22	AA12877	Peptide #1559 enco
30	101.5	5.6	405	22	AA14075	Peptide #2757 enco
31	101.5	5.6	2009	22	AA199674	Human adult form o
32	101	5.5	521	21	AA13578	Streptomyces globi
33	101	5.5	521	21	AA13605	Streptomyces globi
34	100.5	5.5	459	12	AA13498	P. denitrificans CO
35	100.5	5.5	697	22	AA178983	C. glutamicum SRT
36	100.5	5.5	772	22	AA192907	C. glutamicum prote
37	99.5	5.5	412	20	AA14836	Hyperthermostable
38	99.5	5.5	522	18	AA124122	Pyrococcus furiosu
39	99.5	5.5	522	20	AA14838	Hyperthermostable
40	99.5	5.5	654	18	AA124129	Pyrococcus furiosu
41	99.5	5.5	654	20	AA14841	Hyperthermostable
42	99	5.4	1272	21	AA170596	Rat multidrug resi
43	99	5.4	1272	21	AA170597	Rat multidrug resi
44	98.5	5.4	440	20	AA100194	Enterococcus faeca
45	98.5	5.4	2005	22	AA199676	Human adult form o

#### ALIGNMENTS

```

RESULT 1
AAB66708
ID AAB66708 standard; protein; 362 AA.
XX
AC AAB66708;
XX
DT 09-APR-2001 (first entry)
XX
DE C. glutamicum phosphoenolpyruvate protein #2.
XX
KW Phosphoenolpyruvate; sugar phosphotransferase system; PTS.
XX
OS Corynebacterium glutamicum.
XX
PN WO200102583-A2.
XX
PD 11-JAN-2001.
XX
PF 27-JUN-2000; 2000WO-IB00973.
XX
PR 01-JUL-1999; 99US-0142691.
PR 23-AUG-1999; 99US-0150310.
PR 03-SEP-1999; 99DE-1042095.
PR 03-SEP-1999; 99DE-1042097.
XX
(BADI ) BASF AG.
Pompejus M, Kroeger B, Schroeder H, Zelder O, Habershauer G;
WPI; 2001-080989/09.
Corynebacterium glutamicum nucleic acids encoding phosphoenolpyruvate:
sugar phosphotransferase system proteins or their portions, useful for
typing or identifying C. glutamicum or related bacteria, and as markers

```

PT for transformation -

XX  
PS Claim 6; Page 104-106; 144pp; English.  
XX  
XX The present invention relates to Corynebacterium glutamicum  
CC Phosphoenolpyruvate: sugar phosphotransferase system (PTS) proteins.  
CC The PTS nucleic acids and proteins are useful in the  
CC identification of microorganisms which can be used to produce fine  
CC chemicals, for modulating fine chemical production in C. glutamicum or  
CC related bacteria, the typing or identification of C. glutamicum or  
CC related bacteria, as reference points for mapping C. glutamicum genome,  
CC and as markers for transformation..  
XX  
SQ Sequence 362 AA;

Query Match 100.0%; Score 1821; DB 22; Length 362;  
Best Local Similarity 100.0%; Pred. No. 6.5e-171;  
Matches 362; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YDFGGPVGGLFGLVSPVITGLHOSFPPIELEFNQGSFIFATASMANIAQGAACLA 60  
DB 1 YDFGGPVGGLFGLVSPVITGLHOSFPPIELEFNQGSFIFATASMANIAQGAACLA 60  
QY 61 VFELAKSEKLGKLAGASGVSALVIGITEPAIFGVNLRMPFFIGTAAIGGALIALFNI 120  
DB 61 VFELAKSEKLGKLAGASGVSALVIGITEPAIFGVNLRMPFFIGTAAIGGALIALFNI 120  
QY 121 KAVALGAAGFLGVVSDADPMVMFLVCVVTFPIAFGAATAYGLYLVRNNGSIDPDATAA 180  
DB 121 KAVALGAAGFLGVVSDADPMVMFLVCVVTFPIAFGAATAYGLYLVRNNGSIDPDATAA 180  
QY 181 PVPAGTTKAEAEAPAEFSNDSTIIQAPLTGEAIALSSVSDAMFASGKLGSGVAIVPTKGQ 240  
DB 181 PVPAGTTKAEAEAPAEFSNDSTIIQAPLTGEAIALSSVSDAMFASGKLGSGVAIVPTKGQ 240  
QY 241 LVSPVSGKIVWAPPSGHAFVARTKADGNSVDILMHIGFTVNLNGTHFNPLKKQGVDEVK 300  
DB 241 LVSPVSGKIVWAPPSGHAFVARTKADGNSVDILMHIGFTVNLNGTHFNPLKKQGVDEVK 300  
QY 301 AGEELCEFDIDAIAKAAGYEVTPPIVVSNNYKKTGPVNTYGLGTEAGANLLNVAKKEAVPA 360  
DB 301 AGEELCEFDIDAIAKAAGYEVTPPIVVSNNYKKTGPVNTYGLGTEAGANLLNVAKKEAVPA 360  
QY 361 TP 362  
DB 361 TP 362

RESULT 2

AAB66707

ID AAB66707 standard; protein; 468 AA.

XX AAB66707;

AC AAB66707;

XX 09-APR-2001 (first entry)

XX C.glutamicum phosphoenolpyruvate protein #1.

DE Phosphoenolpyruvate; sugar phosphotransferase system; PTS.

XX Corynebacterium glutamicum.

OS WO200102583-A2.

XX 11-JAN-2001.

XX 27-JUN-2000; 2000WO-IB00973.

XX 01-JUL-1999; 99US-0142691.

PR 23-AUG-1999; 99US-0150310.

PR 03-SEP-1999; 99DE-1042095.

PR 03-SEP-1999; 99DE-1042097.

XX (BADI ) BASF AG.

XX Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;

XX WPI; 2001-080989/09.

DR Corynebacterium glutamicum nucleic acids encoding phosphoenolpyruvate;

XX sugar phosphotransferase system proteins or their portions, useful for

PT typing or identifying C. glutamicum or related bacteria, and as markers

PT for transformation -

XX

PS Claim 4; Page 101-102; 144pp; English.

XX The present invention relates to Corynebacterium glutamicum

CC phosphoenolpyruvate: sugar phosphotransferase system (PTS) proteins.

CC The PTS nucleic acids and proteins are useful in the

CC identification of microorganisms which can be used to produce fine

CC chemicals, for modulating fine chemical production in C. glutamicum or

CC related bacteria, the typing or identification of C. glutamicum or

CC related bacteria, as reference points for mapping C. glutamicum genome,

CC and as markers for transformation.

XX

SQ Sequence 468 AA;

Query Match 100.0%; Score 1821; DB 22; Length 468;

Best Local Similarity 100.0%; Pred. No. 9.4e-171;

Matches 362; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YDFGGPVGGLFGLVSPVITGLHOSFPPIELEFNQGSFIFATASMANIAQGAACLA 60

DB 107 YDFGGPVGGLFGLVSPVITGLHOSFPPIELEFNQGSFIFATASMANIAQGAACLA 166

QY 61 VFELAKSEKLGKLAGASGVSALVIGITEPAIFGVNLRMPFFIGTAAIGGALIALFNI 120

DB 167 VFELAKSEKLGKLAGASGVSALVIGITEPAIFGVNLRMPFFIGTAAIGGALIALFNI 226

QY 121 KAVALGAAGFLGVVSDADPMVMFLVCVVTFPIAFGAATAYGLYLVRNNGSIDPDATAA 180

DB 227 KAVALGAAGFLGVVSDADPMVMFLVCVVTFPIAFGAATAYGLYLVRNNGSIDPDATAA 286

QY 181 PVPAGTTKAEAEAPAEFSNDSTIIQAPLTGEAIALSSVSDAMFASGKLGSGVAIVPTKGQ 240

DB 287 PVPAGTTKAEAEAPAEFSNDSTIIQAPLTGEAIALSSVSDAMFASGKLGSGVAIVPTKGQ 346

QY 241 LVSPVSGKIVWAPPSGHAFVARTKADGNSVDILMHIGFTVNLNGTHFNPLKKQGVDEVK 300

DB 347 LVSPVSGKIVWAPPSGHAFVARTKADGNSVDILMHIGFTVNLNGTHFNPLKKQGVDEVK 406

QY 301 AGEELCEFDIDAIAKAAGYEVTPPIVVSNNYKKTGPVNTYGLGTEAGANLLNVAKKEAVPA 360

DB 407 AGEELCEFDIDAIAKAAGYEVTPPIVVSNNYKKTGPVNTYGLGTEAGANLLNVAKKEAVPA 466

QY 361 TP 362

DB 467 TP 468

RESULT 3

AAG92650

ID AAG92650 standard; Protein; 661 AA.

XX AAG92650;

AC AAG92650;

XX 26-SEP-2001 (first entry)

XX C glutamicum protein fragment SEQ ID NO: 6404.

DE Coryneform bacterium; amino acid synthesis; vitamin; saccharide;

XX organic acid synthesis.

XX Corynebacterium glutamicum.



```
XX EP1108790-A2.
PN
XX
XX 20-JUN-2001.
PD
XX
XX 18-DEC-2000; 2000EP-0127688.
PF
XX
XX 16-DEC-1999; 99JP-0377484.
PR
XX
XX 07-APR-2000; 2000JP-0159162.
PR
XX
XX 03-AUG-2000; 2000JP-0280988.
PR
XX
XX (KYOW ) KYOWA HAKKO KOGYO KK.
PA
XX
XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX
XX WPI; 2001-376931/40.
DR
XX
XX N-PSDB; AAH67869.
XX
XX Novel polynucleotides derived from Coryneform bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
PT
XX
XX Claim 17; SEQ ID NO: 6404; 246pp + Sequence Listing; English.
XX
XX The present invention provides a number of nucleotide and protein
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of coryneform bacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from Coryneform bacterium, and identifying a homologue of a gene derived
CC from Coryneform bacterium. Coryneform bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a protein described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
XX
XX Sequence 661 AA;
SQ
Query Match 100.0%; Score 1821; DB 22; Length 661;
Best Local Similarity 100.0%; Pred. No. 1.6e-170; Indels 0; Gaps 0;
Matches 362; Conservative 0; Mismatches 0;
QY 1 YDFGPGVGLLFGLVYSPITVITGLHQSPPIELELFNNGGSPFIFATASMANIAQGAACLA 60
DB 300 ydfgpgvgllfglvyspivltglhqsfpplielelfnnggsgfifatasmanlaqgaacila 359
QY 61 VFFLAKSEKLKGLAGASGVSavlGITEPAIFGVNRLRWPFPIGIGTAIGGALLIALFNI 120
DB 360 vfflakseklkglagsgvsavlgitepaifgvnlrlrwpffigigtaiaggallialfni 419
QY 121 KAVALGAAGFLGVVSDIDPDMVMFLVCVVTFEAFGAIAIYGLVLRNNGSIDPDATAA 180
DB 420 kavalgaagfllgvvsidpdmvmflvcvvtfflfaaiaiyglvlvrnngsidpdataa 479
QY 181 PVPAGTTKAEAPAEFNSNDSTIIQAPLTGEAIALSSVSDAMFASGKLGSGVAIVPTKGQ 240
DB 480 pvpagttkaeapaeafnsndstiiqapltgeaialssvsdamfasklgsgvaivptkgq 539
QY 241 LVSPVSGKIVVAFPSGSHAFVTRKAEKDSNVDILMHIGFDTVNLNGTHFNPLKKGDEVK 300
DB 540 lvspvsgkivvafpsghafvtrkaekdsnvdiilmhigfdtnlngthfnplkkgdevk 599
QY 301 AGECLCEFDIDAIRKAGYEVTTPPIVVSNNKKTGPVNTYGLGEIEAGANLLNVAKEAVPA 360
DB 600 ageclcefdidairkagyevtppivvsnnkktgpvntyglgeieaganllnvakeavpa 659
QY 361 TP 362
XX
```

```
DB 660 tp 661
RESULT 4
AAB69080
ID AAB69080 standard; Protein; 661 AA.
XX
XX AAB69080;
AC
XX
XX 20-APR-2001 (first entry)
DT
XX
XX Brevibacterium lactofermentum sucrose PTS enzyme II protein SEQ ID NO:2.
DE
XX
XX Brevibacterium lactofermentum; sucrose PTS enzyme II; sucrose;
KW phosphoenolpyruvate:carboxylate phosphotransferase system; glucose;
KW coryneform bacterium; phosphoenolpyruvate-sugar transport system.
XX
XX Brevibacterium lactofermentum.
OS
XX
XX WO200102584-A1.
PN
XX
XX 11-JAN-2001.
PD
XX
XX 30-JUN-2000; 2000WO-JP04348.
PF
XX
XX 02-JUL-1999; 99JP-0189512.
PR
XX
XX (AJIN ) AJINOMOTO CO INC.
PA
XX
XX Izui M, Sugimoto M, Nakamatsu T, Kurahashi O;
PI
XX
XX WPI; 2001-138150/14.
DR
XX
XX N-PSDB; AAF32543.
XX
XX Gene encoding sucrose phosphoenolpyruvate-sugar transport system enzyme
PT II obtained by cassette ligation-mediated amplification of downstream
PT domain of coryneform bacterium sucrose gene, with sucrose-binding
PT activity -
XX
XX Claim 1; Page 29-32; 45pp; Japanese.
PS
XX
XX The present sequence represents the Brevibacterium lactofermentum
CC sucrose PTS (phosphoenolpyruvate:carboxylate phosphotransferase system
CC or phosphoenolpyruvate-sugar transport system) enzyme II, which has
CC sucrose-binding activity. A coryneform bacteria produced with the
CC sucrose PTS enzyme II gene can have more efficient sugar uptake, and
CC improved amino-acid and nucleic acid productivity. The sucrose PTS gene
CC and it's disrupted gene, such as one without the sucrose PTS function,
CC can be used to produce new breeds of coryneform bacterial strains to
CC uptake sugar more efficiently e.g. glucose only or and sucrose, and can
CC have improved amino-acid and nucleic acid productivity.
XX
XX Sequence 661 AA;
SQ
Query Match 99.2%; Score 1806; DB 22; Length 661;
Best Local Similarity 99.2%; Pred. No. 4.7e-169;
Matches 359; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 YDFGPGVGLLFGLVYSPITVITGLHQSPPIELELFNNGGSPFIFATASMANIAQGAACLA 60
DB 300 ydfgpgvgllfglvyspivltglhqsfpplielelfnnggsgfifatasmanlaqgaacila 359
QY 61 VFFLAKSEKLKGLAGASGVSavlGITEPAIFGVNRLRWPFPIGIGTAIGGALLIALFNI 120
DB 360 vfflakseklkglagsgvsavlgitepaifgvnlrlrwpfvgigtaiaggallialfni 419
QY 121 KAVALGAAGFLGVVSDIDPDMVMFLVCVVTFEAFGAIAIYGLVLRNNGSIDPDATAA 180
DB 420 kavalgaagfllgvvsidpdmvmflvcvvtfflfaaiaiyglvlvrnngsidpdataa 479
QY 181 PVPAGTTKAEAPAEFNSNDSTIIQAPLTGEAIALSSVSDAMFASGKLGSGVAIVPTKGQ 240
XX
```

Db 480 pvpagttkaeaeapafndstliiqaftgeaialssvsdamfasgklgsqvaivptkqg 539  
QY 241 LVSPVSGKIIVAPSGHAFVATKAEDGSNVDLIMHIGFDTVNLNGTHFNPLKKGQDEVK 300  
Db 540 lvspvsgkiivapsghafavrtkaedgsnvdlimhigfdtnlgtfnplkkgdevk 599  
QY 301 AGELLCEFDIDAIAKAAGYEVTTPIVVSNYKKTGPVNTYGLGEIEAGANLLNVAKKEAVPA 360  
Db 600 agellcefdidaikaagyevttipivsvnykktgpvntyglgeieaganllnvaakeavpa 659  
QY 361 TP 362  
Db 660 tp 661

RESULT 5  
AAG93207  
ID AAG93207 standard; Protein; 683 AA.  
XX  
AC AAG93207;  
DT 26-SEP-2001 (first entry)  
XX  
C glutamicum protein fragment SEQ ID NO: 6961.  
DE  
XX Corynebacterium; amino acid synthesis; vitamin; saccharide;  
KW organic acid synthesis.  
XX  
OS Corynebacterium glutamicum.  
XX  
PN EP1108790-A2.  
XX  
PD 20-JUN-2001.  
XX  
PF 18-DEC-2000; 2000BP-0127688.  
XX  
PR 16-DEC-1999; 99JP-0377484.  
PR 07-APR-2000; 2000JP-0159162.  
PR 03-AUG-2000; 2000JP-0280988.  
PA (KYOW ) KYOWA HAKKO KOGYO KK.  
XX  
XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;  
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;  
DR  
DR WPI: 2001-376931/40.  
DR N-PSDB; AAH68426.  
XX  
PT Novel polynucleotides derived from Coryneform bacteria, for identifying  
PT mutation point of a gene, measuring expression of a gene, analysing  
PT expression profile or pattern of a gene and identifying homologous gene  
PT  
XX  
PS Claim 29; SEQ ID NO: 6961; 246pp + Sequence Listing; English.  
XX  
CC The present invention provides a number of nucleotide and protein  
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These  
CC are useful for identifying the mutation point of a gene derived from a  
CC mutant of coryneform bacterium, measuring expression amount and  
CC analysing the expression profile or expression pattern of a gene derived  
CC from Coryneform bacterium, and identifying a homologue of a gene derived  
CC from coryneform bacterium. Coryneform bacteria are useful for producing  
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,  
CC particularly L-lysine. The present sequence is a protein described  
CC in the exemplification of the invention.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from the  
CC European Patent Office.  
XX  
XX Sequence 683 AA;

Query Match

21.4%; Score 390.5; DB 22; Length 683;

Best Local Similarity 30.7%; Pred. No. 1.3e-29;  
Matches 118; Conservative 60; Mismatches 159; Indels 47; Gaps 9;  
QY 10 LFLGLVYPIVITGLHQSPPIELELFNQ-GGSFIFATASMANIAOGAACLAVFFFLAKSE 68  
Db 307 iviplypflvplghwplnaimlqntlgydfgpgmgawnfacfglvtgfvllsike 366  
QY 69 KLKGLAGAS--GVSAYL--GITEPATFGVNLRLRWPFIFIGTAAIGGALIALFNKAVA 124  
Db 367 rnkamrqvslgmgllagllggiseplsvllrfkkyfrllp9clagglvimglfdikaya 426  
QY 125 LGAAGFLGVVSDIDAPDMVMFLVCAVVTFFIAFGAAAIAYGLYLVRNRGSDID----- 174  
Db 427 fvftslitipamd--pwlgytigiavaffvsmflviald--yrsneerdearakvaadk 481  
QY 175 -----PDAT-AAPV-----PAGTTKAEAEAPESNDSTIIQAPLTGE 211  
Db 482 qaeedikaeanatpaapvaaagagagagagaataavaakpklaagevdivsplegk 541  
QY 212 AIALSSVSDAMFASGKLGSGVAIVPTKGQLVSPVSGKIVVAFPSGHAFVATKAEDGSNV 271  
Db 542 aiplsevpdpifaagklgpgialqptgntvvpadatcvllvqksghavairl-----dsqv 597  
QY 272 DILMHIGFDTVNLNGTHFNPLKKGQDEVKAGELICEFDIDAIAKAAGYEVTTPIVVSNYKK 331  
Db 598 eilvhvgldtvqlggegftvhrqqvkgadplttfdadfirskdplltipvsvsnaak 657  
QY 332 TGPVNTYGLGEIEAGANLLNVAKK 355  
Db 658 fgeiegipadqanssttvikvngk 681

RESULT 6  
AAB66721  
ID AAB66721 standard; protein; 683 AA.  
XX  
AC AAB66721;  
DT 09-APR-2001 (first entry)  
XX  
DE C.glutamicum phosphoenolpyruvate protein #5.  
XX  
KW Phosphoenolpyruvate; sugar phosphotransferase system; PTS.  
XX  
OS Corynebacterium glutamicum.  
XX  
PN WO200102583-A2.  
XX  
PD 11-JAN-2001.  
XX  
PF 27-JUN-2000; 2000WO-IB00973.  
XX  
PR 01-JUL-1999; 99US-0142691.  
PR 23-AUG-1999; 99US-0150310.  
PR 03-SEP-1999; 99DE-1042095.  
PR 03-SEP-1999; 99DE-1042097.  
XX (BADI ) BASF AG.  
XX  
XX Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;  
XX WPI: 2001-080989/09.  
XX  
PT Corynebacterium glutamicum nucleic acids encoding phosphoenolpyruvate:  
PT sugar phosphotransferase system proteins or their portions, useful for  
PT typing or identifying C. glutamicum or related bacteria, and as markers  
PT for transformation  
XX  
PS Claim 32; Page 136; 144pp; English.  
XX  
CC The present invention relates to Corynebacterium glutamicum  
CC phosphoenolpyruvate: sugar phosphotransferase system (PTS) proteins.  
CC The PTS nucleic acids and proteins are useful in the







```

XX AC AAW28051;
XX DT 27-AUG-1998 (first entry)
XX DE Amino acid sequence of Staphylococcus aureus protein.
XX KW Staphylococcus aureus protein; ribozyme; antisense sequence; control;
XX KW Staphylococcal gene; regulatory element; bacterial gene expression;
XX KW vaccine; Staphylococcal infection; food poisoning; scaled skin syndrome;
XX KW toxic shock syndrome.
XX OS Staphylococcus aureus.
XX FH Key Location/Qualifiers
XX FT Misc-difference 1..126 /note= "residues designated X are not defined in
XX FT the specification"
XX PN W09730070-A1.
XX XX 21-AUG-1997.
XX PD
XX XX 19-FEB-1997; 97WO-US02318.
XX XX 20-FEB-1996; 96US-0011888.
XX PA (SMIK ) SMITHKLINE BEECHAM CORP.
XX PI Black MT, Burnham MK, Hodgson JE, Knowles DJC, Nicholas RO;
XX PI Pratt JM, Reichard RW, Rosenberg M, Ward JM;
XX XX
XX DR WPI: 1997-424969/39.
XX DR N-PSDB; AAT84002.
XX XX
XX PT Novel polypeptide(s) from Staphylococcus aureus strain WCUH29 - used
XX PT to isolate antimicrobial compounds, and in vaccines against S.
XX PT aureus infection
XX XX
XX PS Claim 6; Pages 432-433; 989pp; English.
XX CC The present sequence represents a Staphylococcus aureus protein, that,
XX CC based on homology with a human or Staphylococcus aureus protein, is
XX CC believed to be a nerve growth factor or p15g protein. The DNA sequence
XX CC was isolated from a library of clones of S. aureus WCUH 29 in
XX CC Escherichia coli. The DNA sequence can be used in the construction of
XX CC ribozymes and antisense sequences to control the expression of
XX CC Staphylococcal genes. The DNA sequence is also useful as a source of
XX CC regulatory elements for the control of bacterial gene expression. The
XX CC present protein may be used to produce vaccines to enable a host to
XX CC produce specific antibodies with antibacterial action. These vaccines
XX CC and antibodies would protect a host against invasion by S. aureus, and
XX CC conditions relating to Staphylococcal infection, e.g. Staphylococcal
XX CC food poisoning, scaled skin syndrome, and toxic shock syndrome.
XX SQ Sequence 126 AA;

Query Match 8.9%; Score 161.5; DB 18; Length 126;
Best Local Similarity 32.8%; Pred. No. 3.8e-08;
Matches 40; Conservative 25; Mismatches 44; Indels 13; Gaps 4;

OY 228 LGSVVAI-VPTKQLVSPVSGKIVAVPSGHAFVTRKEDGNSVDILMHIGFDTYNLNG 286
Db 2 vgeglaikheeskvlapngflismivptkxhvgi--qsedg--vdvihgkvxvddleg 57
OY 287 THENPLKKGQGVKAGELLCELFDAIKAAQGEVETTPIVVSNKYKTPVNTYGLGEIAG 346
Db 58 kgfkfyvngndhveaqtqlqfagqyiqxgynadxivvis-----nsadlqkvelt 109
OY 347 AN 348
Db 110 mn 111

```

---

```

RESULT 13
AAB40791
ID AAB40791 standard; Protein; 108 AA.
XX
AC AAB40791;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human ORFX ORF555 polypeptide sequence SEQ ID NO:1110.
XX
XX Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
XX KW vulnery; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
XX KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
XX KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
XX KW hypotensive; dermatological; immunosuppressive; antinflammatory;
XX KW antiviral; antibacterial; antifungal; antirheumatic; antichyroid;
XX KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
XX KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
XX KW cholesterol ester storage; systemic lupus erythematosus; infection;
XX KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
XX KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
XX KW thrombosis; contraceptive.
XX OS Homo sapiens.
XX XX
XX PN W0200058473-A2.
XX XX
XX PD 05-OCT-2000.
XX XX
XX PF 31-MAR-2000; 2000WO-US08621.
XX XX
XX PR 31-MAR-1999; 99US-0127607.
XX PR 02-APR-1999; 99US-0127636.
XX PR 05-APR-1999; 99US-0127728.
XX PR 30-MAR-2000; 2000US-0540763.
XX XX
XX PA (CURA-) CURAGEN CORP.
XX XX
XX PI Shinkets RA, Leach M;
XX XX
XX DR WPI: 2000-602362/57.
XX DR N-PSDB; AAC75000.
XX XX
XX PT Novel nucleic acids and peptides derived from open reading frame X,
XX PT useful for treating e.g. cancers, proliferative disorders,
XX PT neurodegenerative disorders and cardiovascular disease -
XX XX
XX PS Claim 11; Page 1038; 5507pp; English.
XX XX
XX CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
XX CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
XX CC sequences have activities such as: cytostatic; hepatotropic; vulnery;
XX CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
XX CC osteopathic; anticonvulsant; antirheumatic; coagulant; vasotropic;
XX CC immunostimulant; cardiant; thrombolytic; immunosuppressive;
XX CC antidiabetic; hypotensive; dermatological; immunosuppressive;
XX CC antinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
XX CC antichyroid; and antianaemic. The sequences can be used for determining
XX CC the presence of or predisposition to, or preventing or treating
XX CC pathological conditions associated with an ORFX-associated disorder. The
XX CC nucleic acids can be used to express ORFX proteins in gene therapy
XX CC vectors. The proteins and nucleic acids may be used to treat cancers,
XX CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
XX CC graft vs host disease, cardiovascular disease, diabetes mellitus,
XX CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
XX CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
XX CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
XX CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
XX CC nocturnal haemoglobinuria, antinflammatory disease; to enhance

```

CC coagulation; to inhibit thrombosis; and as a contraceptive.

XX Sequence 108 AA;

#### Query Match

Best Local Similarity 8.5%; Score 154; DB 21; Length 108;

Matches 32; Conservative 19; Mismatches 43; Indels 4; Gaps 1;

QY 194 PAEFSNDSTIIQAPTGEAIALSSVSDAMFASGKLGSGVAIVPTKGQLVSPVSGKIIVAF 253

Db 14 psslkpsskivaplggllvlpdqdpvfackmvgdgisldpnsnellapvagtvtqlh 73

QY 254 PSCHFAVATKRAEDSGNVDILMHIGFDTVNLNGTHFNP 291

Db 74 nanhaltitp-----egievlnhigldvtmrgdsypp 107

#### RESULT 14

AAV81556

ID AAV81556 standard; Protein; 449 AA.

XX AC AAV81556;

XX DT 24-MAY-2000 (first entry)

XX DE Streptococcus pneumoniae type 4 protein sequence #56.

XX KW Streptococcus pneumoniae; vaccine; screening; protein antigen;

XX KW antibacterial; antiinflammatory; meningitis; infection; diagnosis;

XX KW pneumococcal disease.

XX OS Streptococcus pneumoniae.

XX PN WO200006737-A2.

XX PD 10-FEB-2000.

XX PF 27-JUL-1999; 99WO-GB02451.

XX PR 27-JUL-1998; 98GB-0016337.

XX PR 19-MAR-1999; 99US-0125164.

XX PA (MICR-) MICROBIAL TECHNIKS LTD.

XX PI Gilbert CFG, Hansbro PM;

XX DR WPI; 2000-195300/17.

XX PT New Streptococcal protein, useful as a vaccine, for diagnosis of  
XX PT pneumococcal diseases and for screening agents capable of antagonizing  
XX PT or inhibiting expression of the protein

XX PS Claim 1; Page 78; 108pp; English.

XX CC AAV81501 to AAV81679 represent specifically claimed protein sequences  
XX CC isolated from Streptococcus pneumoniae. AAV81501 to AAV81679 represent  
XX CC specifically claimed nucleotide sequences isolated from S. pneumoniae.

XX CC The sequences have antibacterial and antiinflammatory properties.  
XX CC The protein sequences, and fragments of them, are useful as immunogens  
XX CC and/or antigens. The nucleotide sequences can be used in vaccines and in  
XX CC diagnostic assays. The proteins and nucleotides can be useful for the  
XX CC detection and diagnosis of S. pneumoniae. The protein sequences are also  
XX CC useful for screening an agent capable of antagonizing, inhibiting or  
XX CC interfering with the function or expression of the proteins in which the  
XX CC agent is useful for treatment or prophylaxis of S. pneumoniae infection  
XX CC and meningitis. AAV81591 to AAV81614 represent primers used in the  
XX CC exemplification of the present invention.

XX Sequence 449 AA;

Query Match

8.3%; Score 150.5; DB 21; Length 449;

Best Local Similarity 26.7%; Pred. No. 3e-06;

Matches 56; Conservative 29; Mismatches 68; Indels 57; Gaps 7;

QY 7 VGGLLFGLVYSPVITITGLHQSPPIEIELEFNOGSGFIFATAS---MANIAQGAA----- 57

Db 255 iglmliylltqlwlvgh-----ganlvfafvspialanmaenaagghfav 301

QY 58 -----CLAVFELAKSEKLKGLAGASGSAVLGITEPAIFGV----- 93

Db 302 agefsnmfviagsgatlgcllyiafaskseqikaigrasvvpalfnneplifgipily 361

QY 94 NLRLRPFPIG-IGTAAIGGALIALFNIRKAV-----ALGAAGFLGVVSDAPDMVMF 144

Db 362 npalaipfilapmvtatyyvanslnfkipiaqvptpvgigafgtadira--vlva 419

QY 145 LVCVVVTFIAFGAAIAYGLYLVRNGSID 174

Db 420 lvcataaflvylpfrvydqkivkeegiz 449

#### RESULT 15

AAG90449

ID AAG90449 standard; Protein; 805 AA.

XX AC AAG90449;

XX DT 26-SEP-2001 (first entry)

XX DE C glutamicum protein fragment SEQ ID NO: 4203.

XX KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;  
XX KW organic acid synthesis.

XX OS Corynebacterium glutamicum.

XX PN EP1108790-A2.

XX PD 20-JUN-2001.

XX PF 18-DEC-2000; 2000EP-0127688.

XX PR 16-DEC-1999; 99JP-0377484.

XX PR 07-APR-2000; 2000JP-0159162.

XX PR 03-AUG-2000; 2000JP-0280988.

XX PA (KYOW ) KYOWA HAKKO KOGYO KK.

XX PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;

XX PI Tateishi N, Senoh A, Ikeda M, Ozaki A;

XX DR WPI; 2001-376931/40.

XX DR N-PSDB; AAH65668.

XX PT Novel polynucleotides derived from Coryneform bacteria, for identifying  
XX PT mutation point of a gene, measuring expression of a gene, analysing  
XX PT expression profile or pattern of a gene and identifying homologous gene

XX PS Claim 17; SEQ ID NO: 4203; 246pp + Sequence Listing; English.

XX CC The present invention provides a number of nucleotide and protein  
XX CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These  
XX CC are useful for identifying the mutation point of a gene derived from a  
XX CC mutant of coryneform bacterium, measuring expression amount and  
XX CC analysing the expression profile or expression pattern of a gene derived  
XX CC from Coryneform bacterium, and identifying a homologue of a gene derived  
XX CC from coryneform bacterium. Coryneform bacteria are useful for producing  
XX CC amino acids, nucleic acids, vitamins, saccharides and organic acids,  
XX CC particularly L-lysine. The present sequence is a protein described  
XX CC in the exemplification of the invention.

XX CC Note: The sequence data for this patent did not form part of the printed  
XX CC specification, but was obtained in electronic format directly from the  
XX CC European Patent Office.

```
XX      Query Match
SQ      Sequence      805 AA;
      6.5%; Score 118.5; DB 22; Length 805;
      Best Local Similarity 22.8%; Pred. No. 0.0099;
      Matches 101; Conservative 55; Mismatches 174; Indels 113; Gaps 20;

QY      7 VGGLLFGLVYSPIVI-----TGLHQSF-----PPIELELENQGGSFIFATASMANIAQG- 55
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      174 lggapf-vvaspdalmewnsgrvgefrvatsdpasle-----aasfsdatvvvas-aegh 227

QY      56 -----AACIAVFFFLAKSEKLGKLAGAS-----GVSAVLGITEP 88
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      228 vdkladsylggrdrcyfillaafvavaaavflvvsaysvltgervvrefglirsvgastp 287

QY      89 AIFGVNLRWRPFFIGI-GTAAIG-GALIALFNKIKAVALGAAGF---LGVVSDIDAPDMVM 143
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      288 qilgsvi-----feagilgvvaagfgapaglmaarlinaarfgirvpdivdlpsstm 342

QY      144 FLVCVVTFETAFGAATAYGLYVRRNGSIDPDATAA-----PVPAGTTKAEA 191
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      343 wliagvqvmsviaalpa--vfvscrksavesistpairstspwfgalwlllagivgag 400

QY      192 EAPAEFSNDSTIIQAPLTGEAIALSSVSDAMFASGKLGSGVAIVPTKQLVSPVSGKIVV 251
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      401 mwayeatsdyrgmr-----svalsiaagsalvcalliatavlvplwllhvfrivggtvp 454

QY      252 AFPSGHAFVTRKAEQDSNVVDILM-----HIGFDTV-----NLNGTHFNPLK 293
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      455 tlglglaaakqksrsaallavilagsalsavlhqahighthlvavakmgggtddmmta 514

QY      294 KQDEVKAGELLCFEFDIDATKAAGYEVTTPIVVSNNYKKTGPVNTYGLGEIEAGANLL--- 350
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      515 ldg-eipagmleeissldgvktaipattaveled---sgnfsvlmlae-edgasvmrag 569

QY      351 -----NVAKKEAVPA 360
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      570 dtgagagglvlgrrnsdpqdaypa 592
```

Search completed: March 21, 2002, 16:22:47  
Job time: 223 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 22, 2002, 08:11:36 ; Search time 1607.29 Seconds

(without alignments)

7414.386 Million cell updates/sec

Title: US-09-604-231-3

Perfect score: 1109

Sequence: 1 tatgatattcggcggtccagt.....gttgaacacctgaagtattcgg 1109

Scoring table:

IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:  
1: em\_estfun:\*  
2: em\_esthum:\*  
3: em\_estom:\*  
4: em\_estom:\*  
5: em\_estpl:\*  
6: em\_estba:\*  
7: em\_estro:\*  
8: em\_estov:\*  
9: em\_htc:\*  
10: gb\_est1:\*  
11: gb\_est2:\*  
12: gb\_htc:\*  
13: gb\_gss:\*  
14: em\_gss\_fun:\*  
15: em\_gss\_hum:\*  
16: em\_gss\_inv:\*  
17: em\_gss\_pln:\*  
18: em\_gss\_pro:\*  
19: em\_gss\_rod:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
C 1	45.8	4.1	473	11	T03017
C 2	41.8	3.8	681	13	CNS02EOD
C 3	41.6	3.8	895	13	CNS0071A
C 4	41.4	3.7	477	11	BE997958
C 5	41.4	3.7	559	10	AW574083
C 6	41.4	3.7	601	11	BE997958
C 7	41.4	3.7	627	11	BE998413
C 8	40.8	3.7	884	13	CNS00600
C 9	40.8	3.7	983	13	CNS03CVD
C 10	40.2	3.6	970	13	CNS03H6V
C 11	39.8	3.6	606	11	BE997957
C 12	39.2	3.5	993	13	CNS00LOP

C 13	38.8	3.5	539	10	AI070146
C 14	38.8	3.5	714	10	AW583970
C 15	38.6	3.5	401	10	AI436816
C 16	38.6	3.5	1204	11	BG391833
C 17	38.4	3.5	1101	13	CNS0021D
C 18	38.2	3.4	525	13	CNS025EN
C 19	38.2	3.4	531	10	AW573719
C 20	38.2	3.4	621	11	BE998387
C 21	38.2	3.4	652	10	AI982977
C 22	38.2	3.4	687	10	AW980716
C 23	38.2	3.4	703	10	AW980735
C 24	38.2	3.4	713	11	BG581619
C 25	38.2	3.4	737	11	BG581892
C 26	38.2	3.4	789	11	BG582423
C 27	38.2	3.4	795	11	BG583630
C 28	37.8	3.4	654	10	AI812147
C 29	37.6	3.4	1101	13	CNS00FXE
C 30	37.4	3.4	460	10	AI318021
C 31	37.4	3.4	552	10	BE71200
C 32	37.4	3.4	904	13	CNS03HAO
C 33	37.2	3.4	421	10	AW173566
C 34	37.2	3.4	561	10	AI917973
C 35	37.2	3.4	594	10	AW440291
C 36	37.2	3.4	622	10	AI499080
C 37	37.2	3.4	650	10	AW516253
C 38	37.2	3.4	710	10	AI564728
C 39	37.2	3.4	906	13	CNS02GJN
C 40	37	3.3	1101	13	CNS01602
C 41	36.8	3.3	458	10	AU060899
C 42	36.8	3.3	642	11	BF599151
C 43	36.8	3.3	1159	13	CNS015XR
C 44	36.6	3.3	902	13	CNS006QP
C 45	36.6	3.3	1101	13	CNS017SY

#### ALIGNMENTS

#### RESULT 1

T03017/c  
LOCUS FB20C8 Fetal brain, Strata gene Homo sapiens cDNA clone FB20C8, mRNA  
DEFINITION T03017 473 bp  
ACCESSION T03017  
VERSION T03017.1 GI:314258  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE 1 (bases 1 to 473)  
AUTHORS Khan, A.S., Wilcox, A.S., Polymeropoulos, M.H., Hopkins, J.A., Stevens, T.J., Robinson, M., Orpana, A.K. and Sikela, J.M.  
TITLE Single pass sequencing and physical and genetic mapping of human CDNA  
JOURNAL Nature Genet. 2, 180-185 (1992)  
MEDLINE 94258200  
COMMENT On Sep 21, 1992 this sequence version replaced gi:279156.  
Contact: Sikela JM  
Department of Pharmacology  
University of Colorado Health Sciences Center  
Box C236, 4200 E. 9th Ave, Denver CO 80262-0236  
Tel: 3032708637  
Fax: 3032707097  
Email: nikkit@tally.uchsc.edu.  
Location/Qualifiers  
1. 473  
/organism="Homo sapiens"  
/db\_xref="ATCC (Inhost):86662"  
/db\_xref="GDB:D052785E"  
/db\_xref="taxon:9606"  
/clone="FB20C8"  
/clone\_lib="Fetal brain, Strata gene"

14-FEB-1997  
EST  
cDNA clone FB20C8, mRNA

/notes="Vector: Lambda ZAP II; The FB library (catalog #937201, Stratagene) was constructed by directional cloning and oligo d(T)-priming in Lambda ZAP II phage utilizing mRNA from a 17 week human fetus (total brain)."

BASE COUNT 121 a 124 c 104 g 123 t 1 others  
ORIGIN

Query Match 4.1%; Score 45.8; DB 11; Length 473;  
Best Local Similarity 59.7%; Pred. No. 0.1;  
Matches 77; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 813 ggatatcttgatgcacattgttttcacacagtaaaactcaacgacgacgacactttaacc 872

Db 275 GGAGATCGTGCATATGGTATCGACACCGTAGCGCTGGAAGGTAAGGCTTTAAACG 216

QY 873 gctgaagaacagggcgatgaagtcacaaagcagggagctgctgtgtaattcatattga 932

Db 215 TCTGGTGGAAAGAGGGTGCAGAGTAAGCGCAGGCAACCGATTCTGGAAATGGATCTGGA 156

QY 933 tgccattaa 941

Db 155 TTACCTGAA 147

# RESULT 2

CNS02EOD 681 bp DNA GSS 13-MAY-2000  
LOCUS  
DEFINITION Tetraodon nigroviridis genome survey sequence T7 end of clone 262H14 of library G from Tetraodon nigroviridis, genomic survey sequence.

AL193990

AL193990.1 GI:7832096

GSS: genome survey sequence.

SOURCE Tetraodon nigroviridis.

ORGANISM Tetraodon nigroviridis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon.

1 (bases 1 to 681)

Roest-Crolius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.

Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis

Unpublished

2 (bases 1 to 681)

Roest-Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F., Saurin,W. and Weissenbach,J.

Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence

Unpublished

3 (bases 1 to 681)

Genoscope.

Direct Submission

This sequence (12-APR-2000) to the EMBL/GenBank/DBJ databases

scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at

http://www.genoscope.cns.fr/Tetraodon.

Location/Qualifiers

1. .681

/organism="Tetraodon nigroviridis"

/db\_xref="taxon:99883"

/clone\_lib="262H14"

/note="Genoscope sequence ID : C0AG262DD07LP1-end : T7"

74 a 105 c 170 g 282 t 50 others

BASE COUNT

ORIGIN

Query Match 3.8%; Score 41.8; DB 13; Length 681;  
Best Local Similarity 40.9%; Pred. No. 1.3;  
Matches 139; Conservative 18; Mismatches 183; Indels 0; Gaps 0;

QY 214 ggccttgacaggtcttcagtgctcgcgtgttcttggtattacgagccctgcgattcttc 273

Db 58 GCGGTTGTGTGTTTGTGTTGTTGTCGCGCYGTGTTGTTGTTGTTGTTGTTGTTGTT 117

QY 274 ggtgtgaaccttcgcctgcctgccttccttcattcaggtatcgtaaccgacgattcgg 333

Db 118 GTTGTGTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 177

QY 334 ggccttcttgattgcactctttaaataatcaagcgagttgctgtggcgctgcaggtttcttg 393

Db 178 GTNGTGTCTTNGCCGTNGTGTGTCYGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTY 237

QY 394 ggtgtgttctctattgatgctccagatatggtcatgttcttctgtgtgagttgttacc 453

Db 238 GTCGCGGTCTGYGTGTCYGTGTTTGTGTTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 297

QY 454 ttcttcattcagcattcgccgacgattgcttattgcttcttacttggttcgcccaacggc 513

Db 298 CTTGCTGCTGTCGCCGTGTCGTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 357

QY 514 agcattgatccagatgcaaccgctgctccagtgccctgcag 553

Db 358 GTTGTGTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 397

# RESULT 3

CNS0071A

LOCUS

DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC #

BACR14809 of RPCI-98 library from Drosophila melanogaster (fruit

fly), genomic survey sequence.

AL086286

AL086286.1 GI:4945153

GSS:

SOURCE fruit fly.

ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 895)

Genoscope.

Direct Submission

Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :

BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr

- Web : www.genoscope.cns.fr)

Determination of this BAC-end sequence was carried out as part of a

collaboration with the Berkeley Drosophila Genome Project (BDGP).

The BDGP is constructing a physical map of the Drosophila

melanogaster genome using these BACs. For further information

please see <http://www.fruitfly.org> The BDGP Drosophila

melanogaster BAC library was prepared by Kazutoyo Osoegawa and

Aaron Mammoler in Pieter de Jong's laboratory in the Department of

Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,

NY. The library is named RPCI-98 and was constructed by partial

EcoRI digestion of Drosophila DNA provided by the BDGP from the

isogenic strain Y2; cn bw sp, the same strain used for the BDGP's

p1 and EST libraries. A more detailed description of the library

and how to order individual BAC clones, the entire library, or

filters for hybridization from the BACPAC Resource Center can be

found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

# FEATURES

Location/Qualifiers

1. .895

/organism="Drosophila melanogaster"

/db\_xref="taxon:7227"

/clone\_lib="RPCI-98"

/note="BACR14809"

/note="end : TET3"

124 a 80 c 204 g 179 t 308 others

BASE COUNT

ORIGIN



```

/clone="pGVN-51G2"
/clone_lib="GVN"
/tissue_type="N2-fixing root nodules"
/dev_stage="effective root nodules harvested one month
post inoculation with Sinorhizobium meliloti"
/lab_host="E. coli strain XL0LR"
/notes="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI; cDNA was prepared from polyA+ enriched RNA from
effective root nodules harvested one month post
inoculation with Sinorhizobium meliloti. The cDNA was
directionally ligated into the Uni-ZAP XR vector from
Stratagene and packaged using Gigapack III Gold packaging
extracts. Plasmids containing cDNA inserts were excised
from the recombinant lambda-ZAP phage using Ex-Assist
helper phage and propagated in XL0LR cells."
BASE COUNT      159 a 112 c 112 g 176 t
ORIGIN

```

```

Query Match      3.7%; Score 41.4; DB 10; Length 559;
Best Local Similarity 50.2%; Pred. No. 1.6;
Matches 102; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

QY 828 cattggttcgacagtaaacctcaacgcgcacgtttaaccgcgtgaagcagg 887
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 316 CATTGACTTCACCAAGAGAGCTTCTCATAGCTTCTTAAGCCTAACACGGATCAT 375

QY 888 cgaatgaagtaaacgaggagctgtgtgaattcgatattgatgccattagctgc 947
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 376 TGGTTTTGTCACATGCATCTGAGTTGCATTTACTTTGACAAATCAAGGCATTAA 435

QY 948 aggtatgaagttaaacacgcgcgattgtgttcgaattacaagaacccgacctg 1007
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 436 CAATGAGAAGTACGCGACTTTGTCGTGAGGCTGATTCTGTATGATGATGACATCT 495

QY 1008 cacttacggtttggcgaaattg 1030
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 496 CTTTAAGGACTTTGTCAGACTTG 518

```

```

RESULT 6
BG580528      601 bp      mRNA      EST      11-APR-2001
LOCUS      EST482254 GVN Medicago truncatula cDNA clone pGVN-57H8 5' end, mRNA
DEFINITION      sequence.
ACCESSION      BG580528
VERSION      BG580528.1 GI:13595592
KEYWORDS      EST.
SOURCE      barrel medic.
ORGANISM      Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
Medicago.
REFERENCE      1 (bases 1 to 601)
AUTHORS      Fedorova,M., Pierson,B.L., Samac,D.A., Vance,C.P., Gantt,G.S., Town
              C.D., Van Aken,S., Utterback,T., Cho,J. and Fraser,C.M.
              ESTs from one month old nitrogen-fixing root nodules of Medicago
              truncatula, 2001
TITLE      Unpublished (2001)
JOURNAL      Contact: Carroll P. Vance
COMMENT      Department of Agronomy and Plant Genetics
              University of Minnesota
              411 Borlaug Hall, 1991 Upper Buford Circle, St. Paul, MN 55108 USA
              Tel: 612 625 5715
              Fax: 651-649-5058
              Email: vance004@maroon.tc.umn.edu
              University of Minnesota name: M381580e TIGR sequence name:
              MTCCT40rk More information is available at: http://www.medicago.org
              Seq primer: SKmod (CTA gAA CTA gtg gAT CC).
              Location/Qualifiers
              1. .601
              /organism="Medicago truncatula"

```

```

/cultivar="genotype A17"
/db_xref="taxon:3880"
/clone="pGVN-57H8"
/clone_lib="GVN"
/tissue_type="N2-fixing root nodules"
/dev_stage="effective root nodules harvested one month
post inoculation with Sinorhizobium meliloti"
/lab_host="E. coli strain XL0LR"
/notes="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI; cDNA was prepared from polyA+ enriched RNA from
effective root nodules harvested one month post
inoculation with Sinorhizobium meliloti. The cDNA was
directionally ligated into the Uni-ZAP XR vector from
Stratagene and packaged using Gigapack III Gold packaging
extracts. Plasmids containing cDNA inserts were excised
from the recombinant lambda-ZAP phage using Ex-Assist
helper phage and propagated in XL0LR cells."
BASE COUNT      170 a 121 c 127 g 183 t
ORIGIN

```

```

Query Match      3.7%; Score 41.4; DB 11; Length 601;
Best Local Similarity 50.2%; Pred. No. 1.7;
Matches 102; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

QY 828 cattggttcgacagtaaacctcaacgcgcacgtttaaccgcgtgaagcagg 887
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 307 CATTGACTTCACCAAGAGAGCTTCTCATAGCTTCTTAAGCCTAACACGGATCAT 366

QY 888 cgaatgaagtaaacgaggagctgtgtgaattcgatattgatgccattagctgc 947
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 367 TGGTTTTGTCACATGCATCTGAGTTGCATTTACTTTGACAAATCAAGGCATTAA 426

QY 948 aggtatgaagttaaacacgcgcgattgtgttcgaattacaagaacccgacctg 1007
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 427 CAATGAGAAGTACGCGACTTTGTCGTGAGGCTGATTCTGTATGATGATGACATCT 486

QY 1008 cacttacggtttggcgaaattg 1030
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 487 CTTTAAGGACTTTGTCAGACTTG 509

```

```

RESULT 7
BG998413      627 bp      mRNA      EST      06-OCT-2000
LOCUS      EST430136 GVSN Medicago truncatula cDNA clone pGVSN-9J12, mRNA
DEFINITION      sequence.
ACCESSION      BG998413
VERSION      BG998413.1 GI:10698689
KEYWORDS      EST.
SOURCE      barrel medic.
ORGANISM      Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
Medicago.
REFERENCE      1 (bases 1 to 627)
AUTHORS      Fedorova,M., Pierson,B.L., Samac,D.A., Gantt,J.S., Vance,C.P., Town
              C.D., Bowman,C.L., Craven,M.B., Cho,J. and Fraser,C.M.
              ESTs from senescent nodules of Medicago truncatula
              Unpublished (2000)
TITLE      Contact: Carroll P. Vance
JOURNAL      Department of Agronomy and Plant Genetics
COMMENT      University of Minnesota
              411 Borlaug Hall, 1991 Upper Buford Circle, St. Paul, MN 55108 USA
              Tel: 612 625 5715
              Fax: 651-649-5058
              Email: vance004@maroon.tc.umn.edu
              University of Minnesota name: M272391e TIGR sequence name:
              MTKAP54TK More information is available at:
              http://chryslie.tamu.edu/medicago
              Seq primer: SKmod (CTA gAA CTA gtg gAT CC).
              Location/Qualifiers

```



```

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 983)
AUTHORS Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
        Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
        Saurin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
        Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 983)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
COMMENT This sequence is a single read and was generated as part of a large
        scale clone-end sequencing project of the Tetraodon nigroviridis
        genome. For more information, please take a look at
        http://www.genoscope.cns.fr/Tetraodon.
FEATURES
        Location/Qualifiers
            source
            1..983
            /organism="Tetraodon nigroviridis"
            /db_xref="taxon:99883"
            /clone_lib="G"
            /note="Genoscope sequence ID : COBG015CB04SP1-end :
            PUC-ori"
BASE COUNT 202 a 289 c 274 g 201 t 17 others
ORIGIN

Query Match 3.7%; Score 40.8; DB 13; Length 983;
Best Local Similarity 53.0%; Pred. No. 2.7;
Matches 87; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY 287 tgcgcgtgcgcgtcttcacgtatcggtaccgcagcagtcacgtatcggtggttgatg 346
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 732 GCCGAGACCGGAGCTGGCCCCCGTCCGCGCCGCGCGCTGTGGGGTTTATTACCGT 673

QY 347 cactcttaatacaaggcagttgcgtggcgctgcaggtttcttgggtggttttcta 406
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 672 TGTATTATTATPACATTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 613

QY 407 ttgatctccagataggtcagttcttgggtggtgagcagttgtt 450
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 612 TTGTCGCGTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 569

RESULT 10
CNS03H6V/C 970 bp DNA GSS 17-MAY-2000
LOCUS Tetraodon nigroviridis genome survey sequence T7 end of clone
DEFINITION 026011 of library G from Tetraodon nigroviridis, genomic survey
        sequence.
ACCESSION AL243904
VERSION AL243904.1 GI:7964916
KEYWORDS GSS; genome survey sequence.
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis
        Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
        Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
        Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes;
        Tetraodontidae; Tetraodon.
REFERENCE 1 (bases 1 to 970)
AUTHORS Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
        Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
        Weissenbach,J.
TITLE Characterization and repeat analysis of the compact genome of the
        freshwater pufferfish Tetraodon nigroviridis
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 970)
AUTHORS Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
        Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
        Saurin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
        Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 970)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
COMMENT This sequence is a single read and was generated as part of a large
        scale clone-end sequencing project of the Tetraodon nigroviridis
        genome. For more information, please take a look at
        http://www.genoscope.cns.fr/Tetraodon.
FEATURES
        Location/Qualifiers
            source
            1..970
            /organism="Tetraodon nigroviridis"
            /db_xref="taxon:99883"
            /clone_lib="G"
            /note="Genoscope sequence ID : COBG026AH06LP1-end : T7"
BASE COUNT 311 a 298 c 194 g 152 t 15 others
ORIGIN

Query Match 3.6%; Score 40.2; DB 13; Length 970;
Best Local Similarity 45.9%; Pred. No. 4;
Matches 130; Conservative 2; Mismatches 151; Indels 0; Gaps 0;

QY 219 tgcaggtgttcaggtgtctccgctgtcttcttggtattacggagcctgcgattctcggtgt 278
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 709 TCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 650

QY 279 gaaccttcgcctgcgctgcgttccttcacgtatcggtacgcgagctatcggtggcgc 338
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 649 TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 590

QY 339 ttgattgacactttaataatcaaggcagttgcgttggcgctgcaggtttcttgggtgt 398
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 589 TCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 530

QY 399 tgtttctattgatgtccagataggtcatgttcttgggtgtgtgcagttgttacccttt 458
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 529 TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 470

QY 459 catcgcattcgcgcgagcagattgcttatggcctttacttggtt 501
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 469 TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 427

RESULT 11
BE997957 606 bp mRNA EST 06-OCT-2000
LOCUS EST429680 GVSN Medicago truncatula cDNA clone pGVSN-8B9, mRNA
DEFINITION sequence.
ACCESSION BE997957
VERSION BE997957.1 GI:10698233
KEYWORDS EST.
SOURCE barrel medic.
ORGANISM Medicago truncatula
        Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
        Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
        Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
        Medicago.
REFERENCE 1 (bases 1 to 606)
AUTHORS Fedorova,M., Pierson,B.L., Samac,D.A., Gantt,J.S., Vance,C.P., Town
        ,C.D., Bowman,C.L., Craven,M.B., Cho,J. and Fraser,C.M.
TITLE ESNs from senescent nodules of Medicago truncatula
JOURNAL Unpublished (2000)
COMMENT Contact: Carroll P. Vance
        Department of Agronomy and Plant Genetics
        University of Minnesota
        411 Borlaug Hall, 1991 Upper Buford Circle, St. Paul, MN 55108 USA
        Tel: 612 625 5715
        Fax: 651-649-5058
        Email: vance004@maroon.tc.umn.edu
        University of Minnesota name: M271935e TIGR sequence name:
        MTKAK05TK More information is available at:

```

http://chrysis.tamu.edu/medicago  
Seq primer: SKmod (CTA gAA CTA gtg gAT CC).  
Location/Qualifiers

## FEATURES

source

```
1..606
/organism="Medicago truncatula"
/cultivar="genotype Al7"
/db_xref="taxon:3880"
/clone="pGVSN-889"
/clone_lib="GVSN"
/tissue_type="senescent root nodules"
/dev_stage="mixture of effective nodules from 40 day old
plants harvested 36 hours post shoot removal and nodules
collected from 2 month old plants at mid-pod stage"
/lab_host="E. coli strain SOLR"
/Note="Vector: pBluescript SK +/-; Site_1: EcoRI; Site_2:
XhoI; cDNA was prepared from polyA+ enriched RNA from the
mixture of effective nodules of 40 day old plants
harvested 36 hours post shoot removal and nodules
collected from 2 month old plants at mid-pod stage. The
cDNA was directionally ligated into the Uni-ZAP XR vector
from Stratagene and packaged using Gigapack III Gold
packaging extracts. Plasmids containing cDNA inserts were
excised from the recombinant lambda-ZAP phage using
Ex-Assist helper phage and propagated in SOLR cells."
```

BASE COUNT

173 a 122 c 127 g 184 t

ORIGIN

```
Query Match      3.6%; Score 39.8; DB 11; Length 606;
Best Local Similarity 49.8%; Pred. No. 4.5;
Matches 101; Conservative 0; Mismatches 102; Indels 0; Gaps 0;

QY 828 catttggttcgacacagtagtaaacctcaacggcagcgaacttaacccgctgaagaagcagg 887
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 317 CATTGACTTTCACCAAGAGAGAGCTTCTCATAGCTTCTCTAAGCCTTAACACGGATCATAA 376

QY 888 cgatgaagtaacagcaggagcgtgctgtgaaatcgatattgatgcctaaaggctac 947
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 377 TGGTTTGTGTCATGCTATGATGCTGATGATGATGATGATGATGATGATGATGATGATGAT 436

QY 948 aggttataggtaaccacccgattgtgttttcgaattacaagaacccgacctgataa 1007
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 437 CAATGAAGAAGTGAAGGACCTTTGTCGAGCTGATGATGATGATGATGATGATGATGATGATGAT 496

QY 1008 cacttacggtttggcggaattg 1030
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 497 CTTTAAAGAGTTGTGCAGACTTG 519
```

RESULT 12

```
CNS00LOP/c      993 bp DNA GSS      03-JUN-1999
LOCUS      Drosophila melanogaster genome survey sequence T7 end of BAC:
DEFINITION      BACR23F02 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
```

ACCESSION AL067821

VERSION AL067821.1 GI:4957842

KEYWORDS GSS.

SOURCE fruit fly.

ORGANISM

```
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
Genoscope.
1 (bases 1 to 993)
```

Direct Submission

```
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
```

COMMENT

```
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
```

please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

## FEATURES

source

```
1..993
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPCI-98"
/clone="BACR23F02"
/Note="end : T7"
```

BASE COUNT 379 a 207 c 186 g 151 t 70 others

ORIGIN

```
Query Match      3.5%; Score 39.2; DB 13; Length 993;
Best Local Similarity 48.4%; Pred. No. 7.4;
Matches 104; Conservative 1; Mismatches 110; Indels 0; Gaps 0;

QY 287 gctcggtggcggtcttcacgtatcggtaccgcagctacgtgcggtttgattg 346
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 229 GCTGCCGTTTGTGGCGCTCATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 170

QY 347 cactcttaataatcaagcaggttcggtggcgctcaggtttcttgggtgtttctta 406
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 169 CAGTTCCTGCTGTTGGTGTGCTGTTGTTGGTGTGCTGCTGCTGCTGCTGCTGCTGCTG 110

QY 407 ttgatctccagatggctcatgttcttggtgtgctgagttgttaccttcttcacgc 466
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 109 TTGATGTTCTGCTGCTGCTGTTGTTGTTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 50

QY 467 tcggcgacgagcattgcttatgaccttacttgggt 501
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 49 TTGTTGGCGCTGTTGCTGTTGTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 15
```

RESULT 13

AI070146/c

LOCUS

DEFINITION

UI-R-Y0-lu-g-12-0-UI.s1 UI-R-Y0 Rattus norvegicus cDNA clone

ACCESSION AI070146

VERSION AI070146.1 GI:3396397

KEYWORDS EST.

SOURCE Norway rat.

ORGANISM

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

REFERENCE

1 (bases 1 to 539)

AUTHORS

Bonaldo,M.F., Lennon,G. and Soares,M.B.

TITLE

Normalization and subtraction: two approaches to facilitate gene

discovery

JOURNAL

Genome Res. 6 (9), 791-806 (1996)

MEDLINE

97044477

COMMENT

Contact: Soares, MB

Program for Rat Gene Discovery and Mapping

University of Iowa

451 Eckstein Medical Research Building Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9565

Email: msoares@blue.weeg.uiowa.edu

The sequence tag present in the cDNA between the NotI site and the

oligo-dT track served to identify it as a clone from the normalized

adult eye library. cDNA Library Preparation: M. Fatima Bonaldo,

Ph.D. Clone distribution: clones will be available through Research







**THIS PAGE BLANK (USPTO)**